Blue Waters
2018 Annual Report
Science Highlights

Space Science

CHEMISTRY AND THE EARLY UNIVERSE

Research Challenge

This project focuses on constructing and executing models of the formation of the stellar nurseries of the first stars in the universe, simulating their hydrodynamic and chemical properties to high resolution. Understanding the initial mass function of the first stars will guide the understanding of the sources of chemical elements in the modern universe, as well as help to refine the understanding of the progenitors of gravitational wave events.

Methods & Codes

The project has constructed a cross-domain software package for the efficient solution of chemical species abundances in hydrodynamic simulations, and also implemented a solver for the gravitational potential as defined in spherical coordinates in three dimensions. Codes used include the GAMER-2 simulation code, the Dengo rate construction package, and the Grackle software package.

Results & Impact

These developments will help to contextualize observations from the James Webb Space Telescope. Making these software packages available and accessible will enable scientific inquiry in a number of related domains.

Why Blue Waters

Blue Waters provided the necessary environment to ensure that the project's solver would scale to the size and capacity required for high-resolution studies.

Allocation: PRAC/2700 Knh
PI: Dinshaw Balsara
University of Notre Dame
Space Science

SIMULATING TWO-FLUID MHD TURBULANCE AND DYNAMOS And A NOVEL PARADIGM FOR GEODESIC MESH MHD
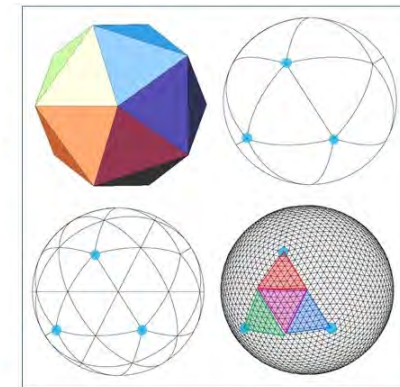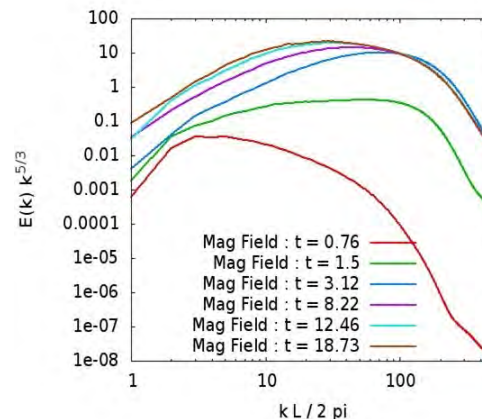
Research Challenge

Our understanding of the star formation process has reached the point where advanced observational capabilities and high-resolution simulations that include the appropriate two-fluid physics of giant molecular clouds (GMCs) are required. We have shown that while magnetic fields grow exponentially in a fully ionized plasma, they grow more slowly in a partially ionized. Another part of our research consists of realizing that most astrophysical systems are spherical and should, therefore, be simulated on meshes that are optimal for such a mesh. We have developed high order MHD for such meshes.
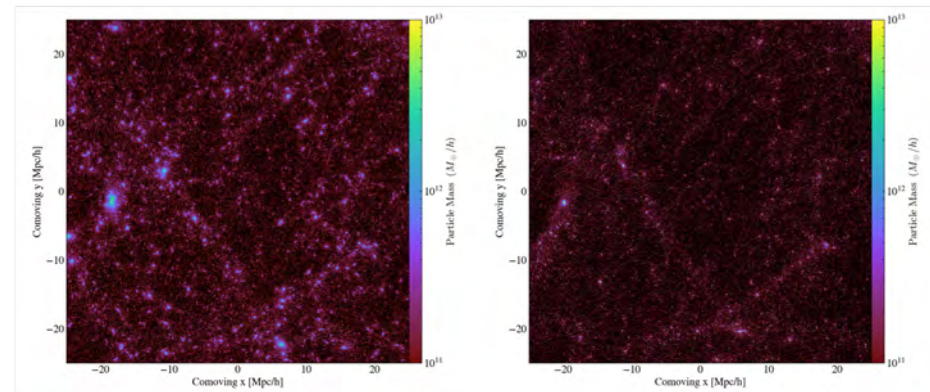

The evolution of magnetic energy with time in a two-fluid dynamo is shown in the left panel. In the right panel we show the mesh subdivision of a geodesic mesh, showing four levels of subdivision. Geodesic mesh MHD is implemented on such a mesh so that the code has no loss of accuracy at the poles, nor does it have a timestep restriction coming from the poles.

Methods & Codes

The core magnetohydrodynamics (MHD) algorithms in our RIEMANN code are based on higher-order Godunov schemes (conservative numerical schemes for solving partial differential equations). The PI and his collaborators have been on the forefront of the effort to develop high-accuracy schemes for computational astrophysics in general and computational MHD in particular, and have previously published results from research in two-fluid methods. We recently published a landmark paper in MNRAS on geodesic mesh MHD.

Why Blue Waters

The simulations reported here are extremely CPU-intensive, and only simulations with very limited resolution (*e.g.,* 512^3 zones) had been done before this Blue Waters PRAC award. The goal of this project requires the petascale computing power of Blue Waters to push the resolution, accuracy, and fidelity of the simulations much higher so we can match theory with the observations coming from NASA-funded instruments. Our codes scale to PetaScale and some of our codes perform extraordinarily well on GPUs. A major technical accomplishment for this year consisted of showing that the geodesic mesh MHD code also scales to PetaScale on BW.

Results & Impact

Several large-scale simulations have been run or are running on Blue Waters, and key results have been published. The present work is a substantial improvement over previous work in terms of resolution and in the details of input physics and accuracy. This new generation of simulations, with $1,024^3$ zones and higher resolution, is able to obtain a well-defined inertial range in two-fluid and dynamo simulations, ensuring that NASA's multi-million dollar investment in observational equipment is being matched by well-resolved simulations. Developing geomesh MHD and publishing it was also a prime accomplishment.

Allocation: Exploratory/50 knh
Matias Carrasco Kind
University of Illinois at Urbana-Champaign
Space Science

THE EVOLUTION OF THE HALO MASS FUNCTION IN AN ACCELERATED UNIVERSE

Research Challenge

The expansion of the universe is dominated by so-called dark energy. By using simulations of the future evolution of dark matter particles, it is possible to place constraints on the bounds of superstructures by following the mass evolution of the biggest clusters found at redshift z = 0 (present day) in the simulation. The challenge is to overcome previous resolution limitations to gain a more complete understanding or how the mass of these structures evolves with time and how that information to put constraints on today's cosmological models of structure formation.

A slice of one of the simulations for the same 50 Mpc/h (megaparsec/hour) region at the present (a=0, left) and in the distant future (a=100, right). Halos are more spherical and well defined in the future; everything is virialized, and no more mass accretion is occurring.

Methods & Codes

- Parallel tree N-body/smoothed particle hydrodynamics (SPH) code GADGET-2
- Simulations started at redshift z = 40 and evolved to scale factor 100, five times older and 100 times larger than the current universe
- Blue Waters' software and packages list: hdf5 (writing the output files), mpich2 (communication), and fftw (tree force computation).

Results & Impact

The team ran several simulations at different scales and are processing and analyzing generated data. Specifically, 100 snapshot per simulations were created to enable evolution of structures in more detail.

They are expecting to publish the results from this simulation in a peer-review journal next year.

Why Blue Waters

N-body dark matter simulations at scale require the use of a high-performance computing resource that is multicore, scalable, stable, and has high bandwidth with low-latency interconnect communication, especially when running large simulations on the order of 1 billion particles. Parallel I/O access to data is also a requirement. Blue Waters provides all of the above in addition to a very helpful dashboard and outstanding help and technical support, which makes it a perfect platform for running such simulations.

SIMULATING THE CO-EVOLUTION OF GALAXIES AND THEIR SUPERMASSIVE BLACK HOLES
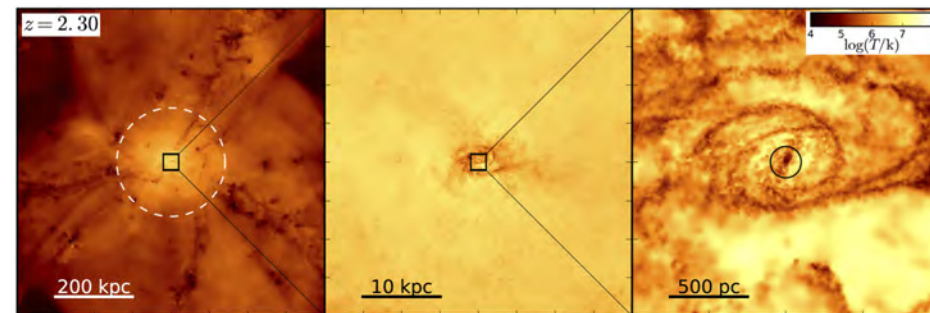
Research Challenge

Observations show a close connection between the growth of supermassive black holes (SMBHs) and the evolution of their host galaxies. Additionally, the history of SMBH growth roughly tracks cosmic star formation history. Also, SMBHs appear important in explaining the properties of the most massive galaxies. The physical mechanisms behind this connection, however, remain largely unknown. The huge, dynamic range of physical scales involved makes it challenging to simultaneously simulate all the relevant dynamics. Key questions remain, including those regarding the origin of galaxy-to-black hole scaling relations, the effects of stellar feedback on SMBH growth, and the triggers and impact of active galaxy nucleus activity (AGN).

Multiscale rendering of the gas distribution (color scale indicates temperature) in one of the simulations of a massive galaxy with a central supermassive black hole (Anglés-Alcázar, et al., 2017). The leftmost panel shows the large-scale distribution of matter (the dashed circle indicates the size of the dark matter halo) and the rightmost panel focuses on the galactic nucleus.

Methods & Codes

To produce galaxy formulation simulations, the researcher team used GIZMO, a publicly available and highly scalable code that implements state-of-the-art gravity and hydrodynamic solvers. The simulations also build on a set of modules for galaxy formation physics developed as part of the FIRE (Feedback In Realistic Environments) project.

Results & Impact

Scientifically, the new simulations pushed the dynamic range of galaxy formation simulations by about an order of magnitude and will enable a wide range of scientific investigations beyond supermassive black holes. Pushing the simulations to the scale of this allocation also enabled researchers to test and optimize the scaling of cosmological zoom-in simulations in a new regime. Since a version of GIZMO is publicly available, this optimization will benefit the broader high-performance computing community.

Why Blue Waters

The Blue Waters computational capabilities are essential to simulate at high resolution the massive galaxies in which most supermassive black hole growth occurs, as these simulations require the largest number of resolution elements.

A contour plot of log density after 20 orbits of evolution subject to the external torque. Overlaid is a line showing the equatorial plane for the spin axis of 24° showing how the inner disk has aligned with the black hole.

ELUCIDATING THE ALIGNMENT MECHANISM FOR BLACK HOLE ACCRETION DISKS SUBJECTED TO LENSE-THIRRING TORQUES
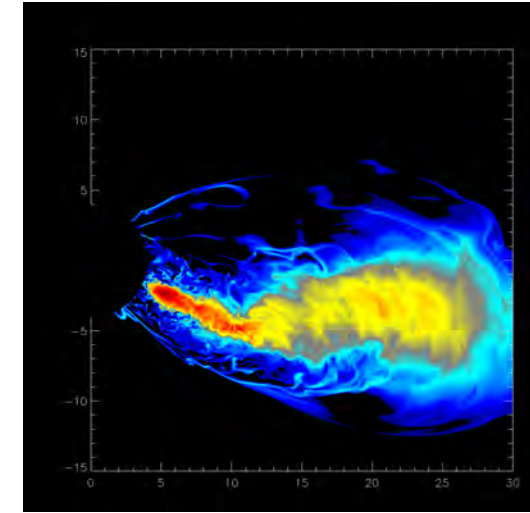
Research Challenge

To study astrophysical accretion onto a spinning black hole in which there is a misalignment between the orbital axis of the incoming gas and the black hole rotation axis. To probe how a time-steady transition occurs between an inner disk region aligned with the equatorial plane of the central mass's spin and an outer region orbiting in a different plane. To utilize the established physical mechanism for internal stresses which requires a timestep very short in comparison to an orbital timescale, whereas the timescale where the orientation transition may occur is likely many orbital periods long.

Methods & Codes

- A simplified disk model consisting of an isothermal disk orbiting a point-mass in Newtonian gravity with a Keplerian angular velocity distribution. It includes only a lowest-order post-Newtonian term to represent the relativistic Lense–Thirring torque.
- Fortran 95 version of *Zeus*, an operator-split code that solves the equations of compressible correlated magnetohydrodynamic turbulence

Results & Impact

- Confirmed that the influence of the sound speed can be encapsulated in a simple "lumped-parameter" model.
- Showed for the range of angles the team studied, the alignment process is largely independent of black hole tilt angles.

Why Blue Waters

The unique high-performance capabilities of Blue Waters enabled key linchpin maximum-resolution simulations that support a wider effort involving a suite of less demanding simulations carried out on other systems.

Allocation: Illinois/224 Knh

PI: Deborah Levin

University of Illinois at Urbana-Champaign

Space Propulsion/Science

CEX ion streamlines (white) superimposed on an ion thruster, ion velocity field. Vertical strips are notional solar cell arrays. Ion impact fluxes differ for two electron models (Boltzmann vs. polytropic.)

MODELING PLASMA FLOWS WITH KINETIC APPROACHES USING HYBRID CPU-GPU COMPUTING

Research Challenge

Ion thruster engines, which are used on spacecraft (satellites), create plasma plumes. Improving prediction of long-term effects of engine emissions on spacecraft surfaces will improve their efficiency and longevity.

Characterize the backflow contamination environment due to the plasma created by electric-propulsion (EP) plumes, and their interaction with the spacecraft environment and neutralizer sources.
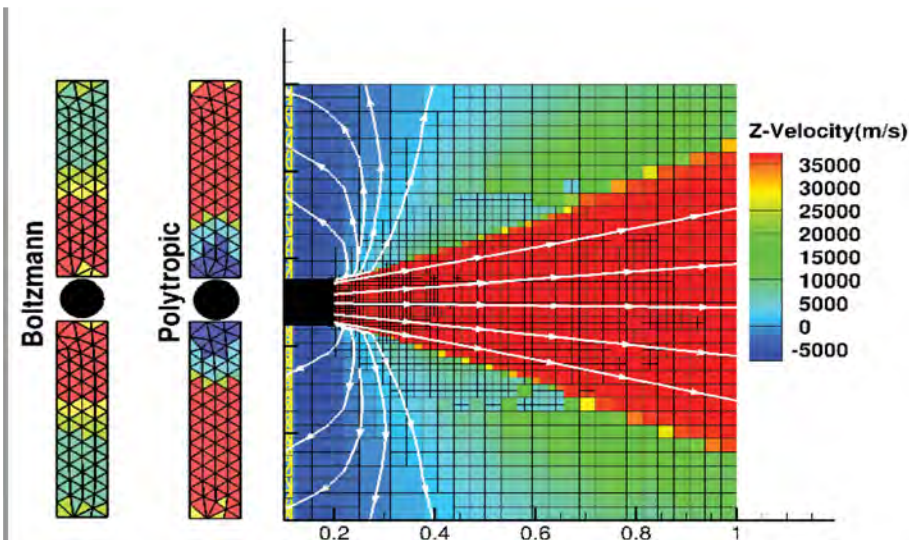
Methods & Codes

Plasma modeling method based on modified DSMC (Direct Simulation Monte Carlo) code, CHAOS (Cuda-based Hybrid Approach to Octree Simulations).

For modeling the electric field, using AMR (Adaptive Mesh Refinement) involved single and multiprocessor stages.

To compute volume of cut-leaf nodes, it utilizes the Morton Z-curve octree structure, a volume-of-fluids (VOF) method, and ray-tracing, which is very efficient on GPUs.

Why Blue Waters

Blue Waters has allowed testing and development of algorithms on a large number of GPUs for three-dimensional, fully kinetic plasma simulations. Compared to the present state-of-the art plasma simulations, a uniform grid in 3D would require a factor of at least ten more cells than the use of AMR/octree. The use of GPUs vs. CPUs decreased the runtime by at least another factor of five.

Results & Impact

High level: Satellites that are more efficient and last longer will save money, as billions of dollars are spent annually building and launching these craft.

Able to model the actual xenon-to-electron mass ratio for three-dimensional geometries (previous modeling of electrons as a separate species in electric-propulsion plumes was limited to two-dimensional cases).

Results support the hypothesis that the ion beam is trapping the electrons, which, in turn, damps the electron oscillations.

DEEP LEARNING FOR MULTIMESSENGER ASTROPHYSICS: REAL TIME DISCOVERY AT SCALE

CNNs can be used for both detection and parameter estimation of GW signals in raw LIGO data.
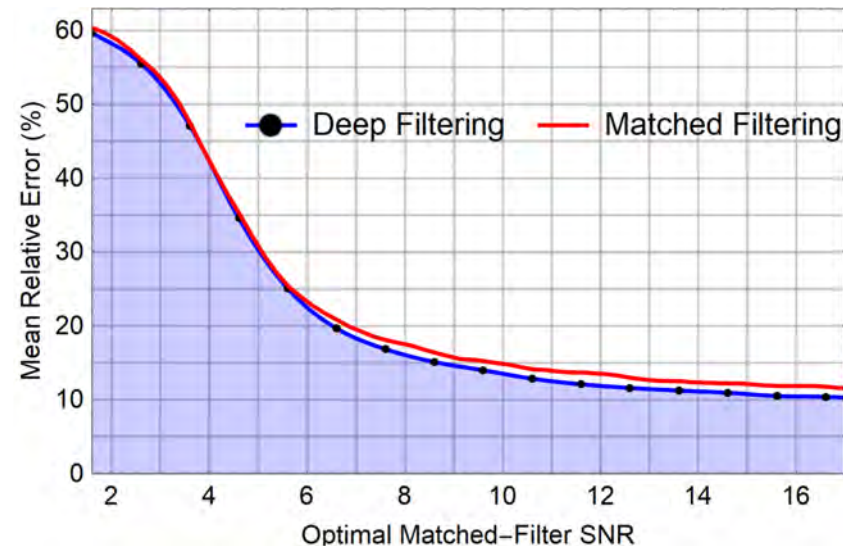
Research Challenge

Matched-filtering searches, the most sensitive gravitational wave (GW) detection algorithms used by LIGO, currently target a 4D parameter. Extending these template-matching searches to target the 9D parameter space available to GW detectors is computationally prohibitive. To address these limitations, we pioneered the use of GPU-accelerated deep learning algorithms that employ a system of two deep convolution neural networks (CNNs) that take time-series inputs for both classification and regression.

Methods & Codes

We used the Wolfram Language neural network functionality, built using the open-source MXNet framework, that uses the cuDNN library for accelerating training on NVIDIA GPUs. The learning algorithm was ADAM. While training, we used the curriculum learning strategy to improve the performance and reduce training times of the CNNs while retaining performance at very high signal-to-noise ratios.

Results & Impact

Deep Filtering demonstrated, for the first time, that machine learning can detect and recover true parameters of real GW signals observed by LIGO and achieve performance comparable to matched-filtering methods, several orders of magnitude faster and far more resilient to transient noise artifacts. We showed that after a single training process, Deep Filtering can automatically generalize to noise having new Power Spectral Densities (PSDs) from different LIGO events, without retraining.

Why Blue Waters

Blue Waters played a critical role in creating the numerical relativity waveforms used to train and test deep learning algorithms. In recent developments, Blue Waters has provided the required scale and computational power to construct deep neural networks using distributed learning involving over 1,024 GPUs.

Allocation: Illinois/230 Knh

PI: Hsi-Yu Schive

University of Illinois at Urbana-Champaign

Space Science

ASTROPHYSICS ON GRAPHICAL PROCESSING UNITS

Simulation of galaxy cluster merger, reaching a ten times higher resolution by utilizing GAMER on Blue Waters.
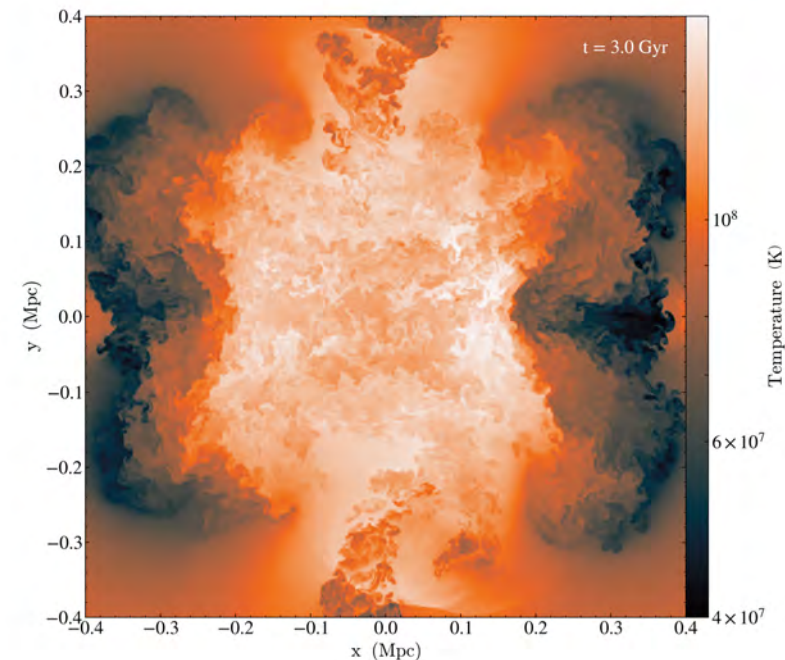
Research Challenge

The computational frameworks used for Astrophysics that utilize the numeric analysis technique of adaptive mesh refinement (AMR), have difficulty achieving an optimum performance on supercomputers, like Blue Waters, that have both GPU and CPUs. This often ends up with an imbalance with the GPU-related components of a job waiting for the CPU parts of a job to finish. Therefore the goal of this project is to address these issues and study astrophysical phenomena requiring resolutions not realistically attainable previously.

Methods & Codes

GAMER is a GPU-accelerated adaptive mesh refinement (AMR) code for astrophysics. Utilizing GAMER, the project has significantly revised the code to incorporate a rich set of new physics and optimized the overall performance and parallel scalability.

Results & Impact

The project has incorporated a rich set of new physics and functionality into the code. To demonstrate the capability of GAMER, it was compared with two widely adopted codes named "Flash" and "Enzo" on realistic astrophysical applications, i.e., galaxy cluster merger and star formation in an isolated disk galaxy. This comparison showed that GAMER not only produces physical results that are very consistent with Flash and Enzo but also runs significantly faster, by one to two orders of magnitude.

Why Blue Waters

Blue Waters provides a unique opportunity for to test and optimize GAMER on an extreme scale using thousands of GPUs and tens of thousands of CPU cores simultaneously. Moreover, the Blue Waters staff is very efficient in helping solve technical issues such as fine-tuning the hybrid OpenMP/MPI model. Without their help, it would have been very difficult to achieve an optimal throughput.
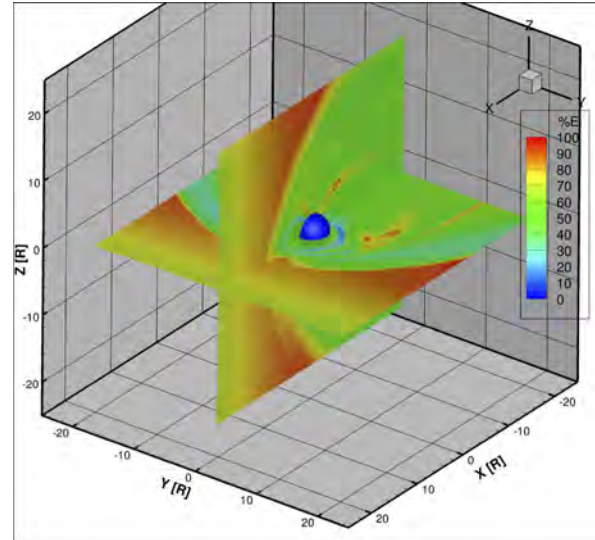
Allocation: Illinois/60 Knh

PI: Raluca Ilie

University of Illinois at Urbana-Champaign

Space Science

The figure shows the percent of the inductive field (color) from the total electric field at a specific time during a geomagnetic storm. The blue sphere represents the inner boundary of the simulation, which is set at 2.5 Earth radii.

INDUCTIVE E FIELDS IN EARTH'S MAGNETOSPHERE

Research Challenge

The overarching goal of this research is to determine what role the electric field induced by the time-varying magnetic field plays in the overall particle energization and to better understand the particle transport across the Earth's magnetosphere. This new knowledge will allow work toward improvements of physics-based space weather prediction models.

Methods & Codes

Due to large variability in both space and time, modeling the Earth's magnetosphere is a very challenging problem. This project uses the Space Weather Modeling Framework (SWMF), which is capable of simulating physical processes and coupled domains from the solar surface to the upper atmosphere of the Earth. The SWMF is a robust, high-performance numerical tool for heliophysical simulations.

Results & Impact

In short, the results suggest that the electric field induced by the time-varying magnetic field plays a crucial role in the overall particle energization in the inner magnetosphere. This highlights the importance of accounting for inductive electric fields in space weather prediction models, a component long ignored in the description of near-Earth plasma dynamics. The implications of these findings are immediate as space weather prediction is critical to a forewarning of solar events that could generate severe space weather at Earth.

Why Blue Waters

This project utilizes an array of numerical approaches based on satellite and ground-based observation. The calculations require high-accuracy, high-resolution simulations across a very large domain, which is not possible on smaller clusters. This involves models consisting of many data points across space and time and requires a large computational resource, such as Blue Waters.
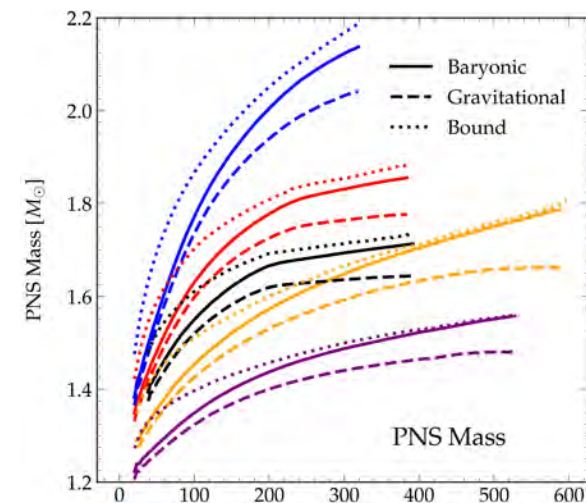
The mass of the neutron star forming at the center of the supernova for the different progenitor stars simulated by the team The highest mass simulation (in blue) is still accreting at 0.45 solar masses per second and will likely collapse to a black hole.

CORE-COLLAPSE SUPERNOVA SIMULATIONS: SIMULATING THE BRIGHTEST OBJECTS IN THE SKY AND THE SOURCE OF LIFE'S BUILDING BLOCKS

Research Challenge

To understand how differing stellar properties, just prior to core collapse, affect the behavior of supernovae events and to characterize the radiation that will be generated from the next supernova to occur in our galaxy.

Methods & Codes

These simulations use the Zelmani Core Collapse Simulation Package and leverages the Cactus Framework. Zelmani provides OpenMP+MPI parallelized adaptive mesh refinement, high-order shock capturing finite volume methods for the fluid and finite difference methods for the metric evolution equations, and handles neutrino radiation using an approximate M1 scheme. All simulation output uses HDF5 for data and metadata storage.

Results & Impact

The team has published a first set of results in [1] where they present a study of the progenitor dependence of a three-dimensional neutrino mechanism of core-collapse supernovae. The results suggest a complex, nonmonotonic dependence on progenitor parameters, hinting at a complex interplay between multiple proposed explosion mechanisms, necessitating more detailed numerical studies to fully understand the effects.

Ott, C.D., et al., The Progenitor Dependence of Core-collapse Supernovae from Three-dimensional Simulations with Progenitor Models of 12–40 M$\odot$. Astrophys.J.,855 (2018). DOI: 10.3847/2041-8213/aaa967

Why Blue Waters

Fully three-dimensional, general-relativistic, radiation-magnetohydrodynamic simulations are simply too demanding of computational resources for any but a leadership-class facility to support them. The project's simulations required the use of hundreds of compute nodes to provide sufficient memory for the simulation's state vector. The exceptional speed of Blue Waters' network provided the capability to scale to the number of nodes required to complete these simulations.
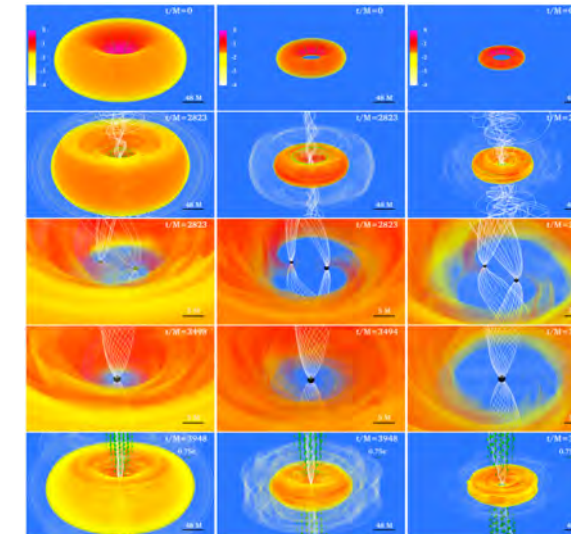
Allocation: Illinois/1499 Knh

PI: Stuart Shapiro

University of Illinois at Urbana-Champaign

Space Science

DISKS AROUND MERGING BINARY BLACK HOLES: FROM GW150914 TO SUPERMASSIVE BLACK HOLES

Research Challenge

The coincident detection of gravitational waves (GWs) with electromagnetic (EM) signals from the coalescence of black hole binaries is new observational challenge. Combining GW and EM observations offers a unique probe to understanding black hole cosmological evolution and accretion processes. We report results from general relativity simulations of circumbinary magnetized disks accreting onto nonspinning merging black holes. We survey different disk models to quantify the robustness of previous simulations on the initial disk model. Scaling our simulations to supermassive binary black holes, we find that the observable flow properties such as accretion rate periodicities and the emergence of jets throughout inspiral, merger, and post merger we reported in earlier studies display only modest dependence on the initial disk model.

Volume rendering of rest-mass density, normalized to its initial maximum value $\rho_{0,max}$ (see color coding), magnetic field lines (solid white curves), and velocity vectors (green arrows) at select times during the inspiral, merger and post-merger. Case A corresponds to the left column, case B to the middle column, and case C to the right column.

Methods & Codes

Magnetohydrodynamic (MHD) numerical simulations in full general relativity (GR) require the solution of the Einstein field equations to determine the gravitational field, the relativistic MHD equations to determine the flow of matter, and the electromagnetic fields. Together, the equations constitute a large system of highly nonlinear, multidimensional, partial differential equations in space and time. We solve these equations through independently developed code, "Illinois GRMHD", which has been built over many years on the Cactus infrastructure and uses the Carpet code for adaptive mesh refinement but employs original algorithms and coding.

Results & Impact

We performed MHD simulations of binary black holes with different mass ratios that accrete magnetized matter from a circumbinary accretion disk. We considered three initial disk models that differ in their scale heights, physical extent, and in their magnetic field content in order to test whether previous properties of MHD accretion flows onto binary black holes are sensitive to the initial disk model. Scaling the simulations to LIGO GW150914 black hole collision we find that magnetized disk accretion onto binary black holes could explain both the GWs detected from this system and the EM counterpart GW150915-GBM reported by the Fermi GBM team 0.4seconds after LIGO's GW150915. When scaling to supermassive black hole binaries, we find that at late times flow properties, temperatures, and thermal frequencies are all robust, displaying only modest dependence on the disk model.

Why Blue Waters

Blue Water provides the required computational power to simulate these cosmic sources in a timely manner. By adding OpenMP support to our MPI-based code, scalability on multicore machines has improved greatly. With the Blue Waters next-generation interconnect and processors, our hybrid OpenMP/MPI code exhibits greater scalability and performance than on any other supercomputer we have used. Recently, we were able to build our code with the Blue Waters Intel compilers. This resulted in a significant boost of our code's performance by about 30%, making Blue Waters unique for tackling the astrophysical problems we want to address.
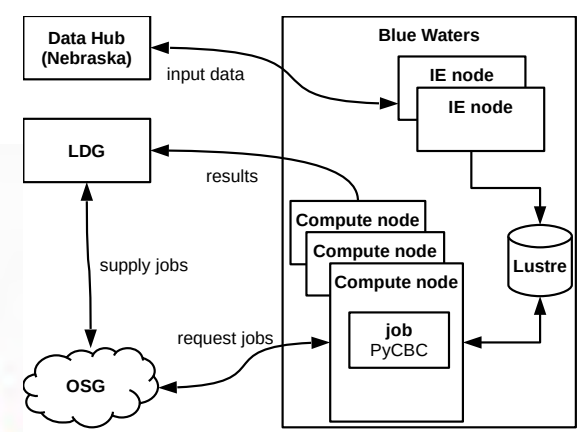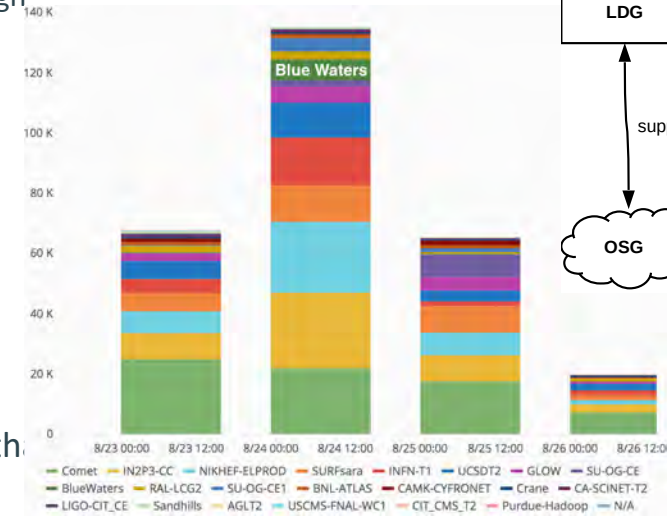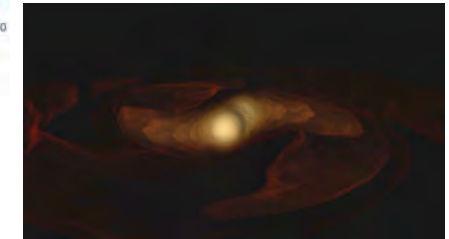
Allocation: Innovation & Exploration/1000 knh

Eliu Huerta

University of Illinois at Urbana-Champaign

Space Science

Integration of NSF cyberinfrastructure

Convergence of large scale workflows, simulation and visualization

MULTIMESSENGER ASTROPHYSICS WITH THE BLUE WATERS SUPERCOMPUTER

Research Challenge

Future gravitational wave discovery campaigns will be longer and will involve more detectors, thereby requiring additional compute resources, including one outside of LIGO's normal processing infrastructure to validate the results in a reasonable timeframe. Multi-Messenger Astrophysics (MMA) will require the interoperability of NSF cyberinfrastructure resources so that large projects can benefit from making use of existing resources rather than building custom solutions for computational workflows. Additional challenges include increasing cluster utilization and throughput on current and future NSF resources.

Methods & Codes

The LIGO workload used in this work is a Python based software stack. Pegasus is used as the workflow management system. LIGO's software was containerized on Blue Waters using Shifter. We configured Blue Waters as an Open Science Grid compute element, and then connect it to the LIGO Data Grid using the Open Science Grid as a universal adapter.

Results & Impact

This works marks the first time convergence was reached on the Blue Waters supercomputer and exhibited the flexibility and interoperability of NSF cyberinfrastructure to enable and accelerate scientific discovery at scale.

We have used this novel computational framework to validate the gravitational wave detection of two colliding neutron stars with the LIGO and Virgo detectors on demand and at scale.

Why Blue Waters

The Blue Waters supercomputer is ideally suited to facilitate large-scale gravitational wave data analysis because the large number of independent jobs matches well with the reasonably large set of otherwise unoccupied nodes available through single-node backfill. BW staff helped us get containers working on the system so that the workflow could be used on both BW and OSG resources at the appropriate scale for each service without rework of the basic code.

Allocation: NSF PRAC/5220 Knh
PI: Alexander Tchekhovskoy
Northwestern University
Space Science

FEEDING BLACK HOLES: TILT WITH A TWIST
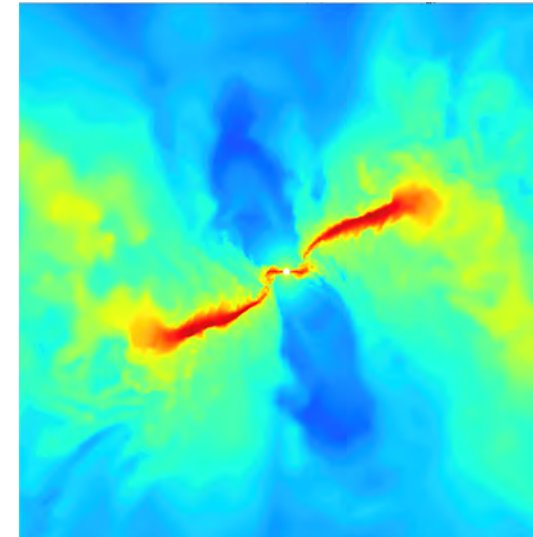
The inner part of a disk of half-thickness h/r = 0.03 tilted by 45 degrees aligns with the black hole midplane and breaks off from the outer misaligned part of the disk. This is the first demonstration of the Bardeen–Petterson effect in a general relativistic magnetohydrodynamic (GRMHD) numerical simulation.

Research Challenge

Studies of accretion disks are extremely challenging, especially in the crucial regime of luminous accretion that powers quasars. Such accretion disks are thin, often tilted relative to the black hole rotation axis, and difficult to resolve numerically, requiring high resolutions and adaptive grids to follow the body of the disk as it moves. For the past 40 years, the inner parts of such disks were expected to align with the black hole. In nature, the disks are magnetized and turbulent, yet no simulation of such a tilted magnetized turbulent disk has shown this long-sought alignment.

Methods & Codes

- Developed the new code H-AMR (pronounced "hammer"), which includes adaptive mesh refinement, local adaptive timestepping, and efficiently runs on GPUs.
- H-AMR is parallelized via MPI with domain decomposition and scales well to thousands of GPUs, achieving weak scaling efficiency of 85 per cent on 4096 GPUs on the Blue Waters

Results & Impact

More than 40 years after it was first proposed by Bardeen and Petterson, this project has finally demonstrated the existence of the Bardeen–Petterson alignment of geometrically thin turbulent magnetized disks. In fact, at larger tilt angles, the inner misaligned part of the disk breaks off from the outer misaligned part.

Why Blue Waters

Blue Waters access has been instrumental to this team's ability to obtain these groundbreaking results, which require not only enormous amounts of computing power but also fast interconnect to make use of hundreds of XK nodes. Further, Blue Waters staff helped enormously with 3D visualization.

Allocation: NSF PRAC/4000 Knh
PI: Brian O'Shea
Michigan State University
Space Science

SIMULATING GALAXY FORMATION ACROSS COSMIC TIME
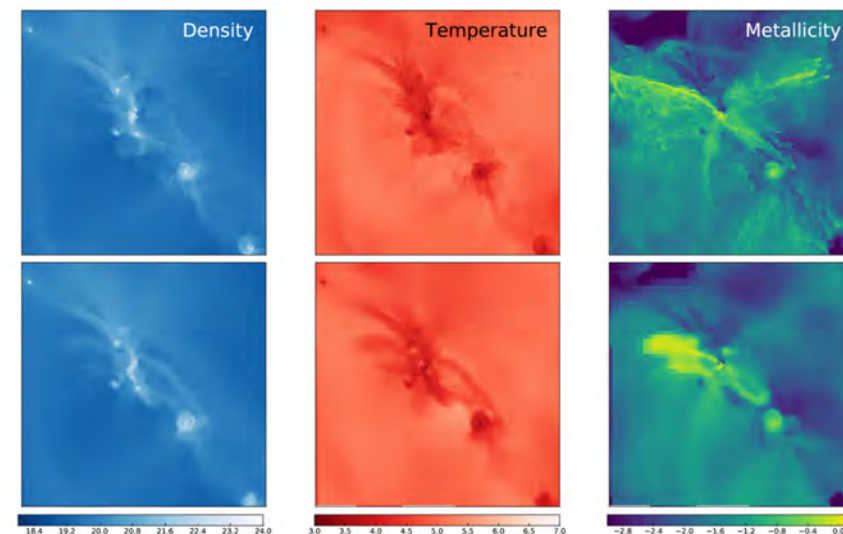
This figure shows a projection of (left to right) gas density, temperature, and metal content for two simulations: one with an enhanced spatial refinement in the circumgalactic gas (top row), and one using standard refinement methods (bottom row).

Research Challenge

To understand two critical issues in galaxy formation:

1. the formation of the earliest generations of galaxies and their connections to the Milky Way through hierarchical structure formation,
2. the "baryon cycle" in galaxies like the Milky Way—how gas gets into and out of galaxies, and what it does while it is there.

Methods & Codes

Enzo is an open-source and community-developed software platform for studying cosmological structure formation. It allows for inclusions of all of the critical physical components needed to study galaxy formation—gravity, dark matter dynamics, fluid dynamics, the microphysics of plasmas, and prescriptions for star formation and feedback—and to do so using a tool that can scale to large numbers of CPUs. All analysis was done with the program **yt**.
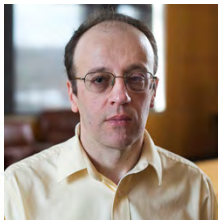
Why Blue Waters

Blue Waters is the only machine available to the academic community that fits all of the requirements for this research, which involves simulations to properly model galaxies in both the early universe and the present day: large memory and disk space, high bandwidth and low-latency communication.

Results & Impact

- While stellar-mass black holes are not capable of growing into billion-solar-mass objects by the time that they can be observed (a billion years after the Big Bang), it is possible for massive gas clouds to directly collapse into much more massive objects that can easily seed black holes.
- Simulations show that increased physical resolution in the circumgalactic medium is incredibly important. Increasing the resolution by more than an order of magnitude beyond previous state-of-the-art calculations resulted in the appearance of both spatial and chemical features that are seen in observations but not in previous models.

MULTISCALE SPACE WEATHER SIMULATIONS

MHD–EPIC simulation of the Earth's magnetosphere. Left: color image represents pressure with projected field lines. The MHD simulation extends to 256Re; the red box indicates the embedded PIC model (4Re × 6Re × 6Re). Right: simulated flux transfer event where a magnetic flux rope is formed by magnetic reconnection. Field line colors represent plasma pressure.
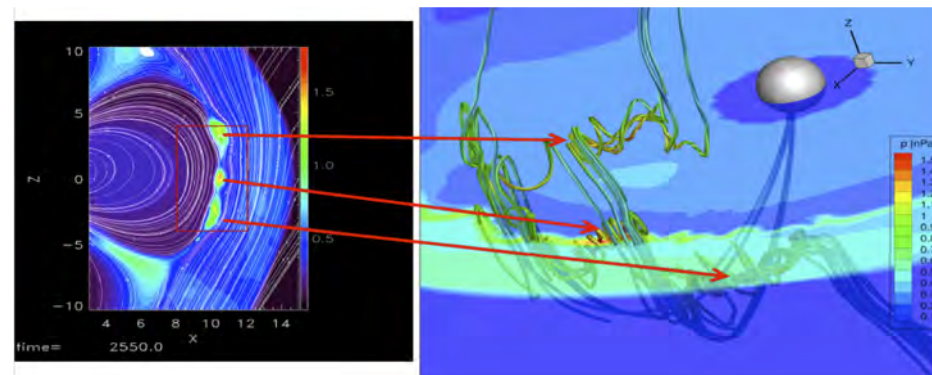
Research Challenge

The research solves fundamental problems in plasma physics, solar physics, and magnetospheric physics that relate to magnetic field energization and reconnection. Consequently, the results are significant to solar and plasma scientists, as well as magnetosphere and space weather scientists.

Methods & Codes

The research team approach combines the efficiency of global fluid-type models with the local kinetic models. The resulting magnetohydrodynamic with embedded particle-in-cell (MHD–EPIC) model is 100 to 10,000 times more efficient than a global kinetic model. The flux emergence and CME initiation simulations are carried out with the high-resolution MHD code BATS–R–US in a configuration called the Spherical Wedge Active Region Model (SWARM). Using SWARM, they performed rigorous flux-emergence calculations and the formation of active regions with no *ad hoc* assumptions about coronal or photospheric conditions.

Results & Impact

Using the kinetic scaling in combination with the MHD-EPIC method , the research team was able to perform the very first three-dimensional global Earth magnetosphere simulations with an embedded kinetic model. Currently they run simulations where the reconnection process of the magnetotail is covered by the kinetic model. The preliminary simulations show that the model can produce the observed dynamics and suggest that the model can capture the interplay of global magnetospheric and local kinetic processes.

Why Blue Waters

The team used the Blue Waters resources to study how the global and local plasma scales are related and how the numerical simulations can be sped up. Despite the algorithmic advances that sped up the simulations by several orders of magnitude, they still require the computational capabilities that are only available at the largest systems, like Blue Waters.
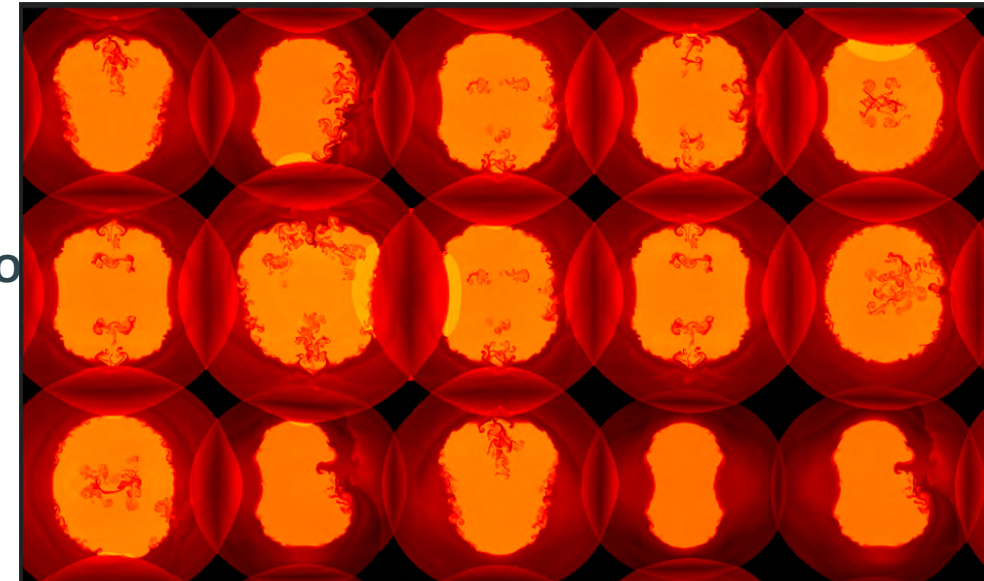
Allocation: NSF PRAC/125 Knh

PI: Michael Norman

University of California, San Diego

Space Science

REALISTIC SIMULATIONS OF THE INTERGALACTIC MEDIUM: THE NEED FOR ENZO-P/CELLO

Research Challenge

Improved computational models of the intergalactic medium (IGM) are needed to extract information encoded in the high-resolution optical spectra of distant quasars. Using the Enzo hydrodynamic cosmology code enhanced with multigroup flux-limited diffusion radiative transfer we found this change did not improve the results. We are therefore investigating our next hypothesis: that **dense gas bound to galaxies that is unresolved in the Enzo simulations supplies significant absorption of the quasar light and modifies the key observables in such a way to improve agreement with observations**.

Methods & Codes

Including galaxies in simulations of the IGM poses **severe resolution requirements** that can be addressed using **adaptive mesh refinement** (AMR). Enzo-P (for petascale) is built on an entirely **new highly scalable AMR framework called Cello**. We have implemented the already proven scalable Forest-of-Octrees AMR algorithm on top of Charm++ and have obtained **excellent parallel scaling results on Blue Waters** as a prelude to our target application problem

Results & Impact

Achieved **ideal** weak scaling to **262k cores** on a hydrodynamic AMR test problem shown above involving an array of blast waves driven by high-pressure regions. The largest problem evolved **1.7 trillion cells**.

This work is significant because it provides a path to **exascale** for the entire Enzo community of over 100 developers and users.
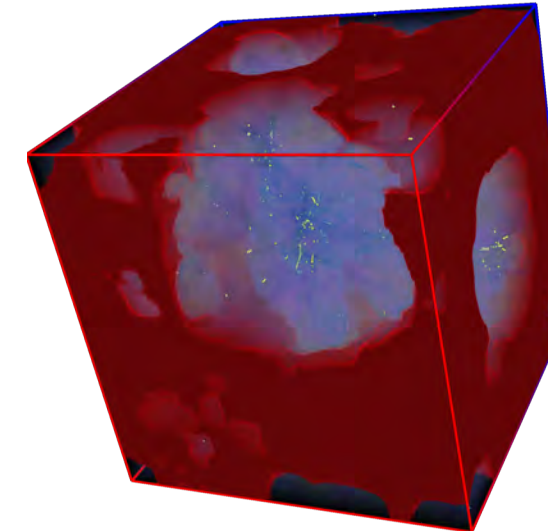
Why Blue Waters

The scale of Blue Waters and the ability to run large scaling tests approaching full system size quickly and reliably has substantially accelerated our code development.

COSMIC REIONIZATION ON COMPUTERS

Propagation of ionized bubbles (transclucent blue color) into the primordial neutral gas (nontransparent dark red). Ionization is driven by early galaxies (yellow dots). The cube is a small region of the whole computational volume, which is 20 times larger in each direction.

Research Challenge

The Cosmic Reionization on Computers (CROC) project aims, over the course of several years, to produce numerical simulations of cosmic reionization that model self-consistently all relevant physics, ranging from radiative transfer to gas dynamics and star formation, in simulation volumes of over 100 comoving megaparsecs (Mpc, necessary to model a representative sample of high-mass galaxies) and with spatial resolution approaching 100 parsecs in physical units (which is necessary to reliably model star formation in galaxies).

Methods & Codes

In order to reach the required dynamic range, the Adaptive Mesh Refinement (AMR) technique was used. The simulations were run with the Adaptive Refinement Tree (ART) code, which includes all necessary physical modules for simulating cosmic reionization (dynamics of dark matter and gas, atomic processes, interstellar chemistry, star formation and stellar feedback, radiative transfer of ionizing and UV radiation).

Results & Impact

Simulations on Blue Waters reach scales of 120 Mpc and are the only existing numerical simulations of cosmic reionization that are not only able to match all of the currently existing observational data, but also achieve sufficient numerical precision and physical fidelity to serve as theoretical counterparts for the soon-to-be-launched James Webb Space Telescope and next generation large ground-bases telescopes.

Why Blue Waters

Blue Waters is the only existing U.S. supercomputer where the project can be efficiently completed. ART code does not support GPUs, so Titan is not suitable for our purposes. Prior to the Blue Waters allocation, the project was supported by the U.S. Department of Energy's INCITE program with a more modest allocation on Mira at the Argonne Leadership Computing Facility. However, Mira is significantly slower than Blue Waters so using Blue Waters is almost 10 times more efficient for these simulations.

Allocation: NSF PRAC/2140 Knh

PI: Nikolai Pogorelov

University of Alabama, Huntsville

Space Science
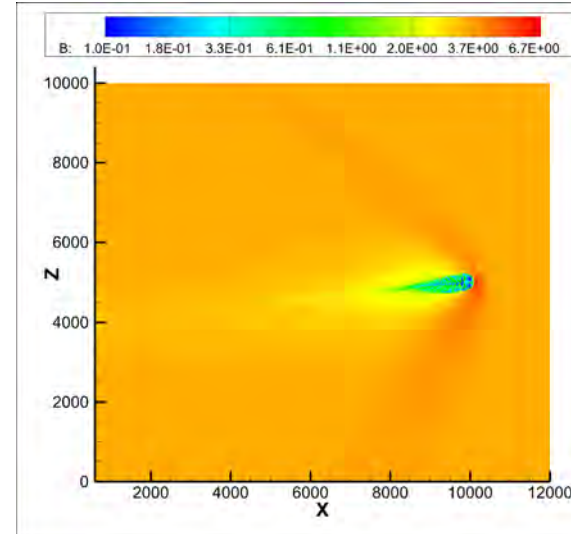
The distribution of magnetic field strength in the plane determined by the Sun's rotation axis and the LISM velocity vector.

MODELING PHYSICAL PROCESSES IN THE SOLAR WIND AND LOCAL INTERSTELLAR MEDIUM WITH A MULTISCALE FLUID-KINETIC SIMULATION SUITE: CONNECTING SOLAR, HELIOSPHERIC, AND ASTROPHYSICAL SCALES

Research Challenge

The grand challenge of this research is to create computer simulations of the interactions of solar wind within the solar system as well as with the Local Interstellar Medium (outside the solar system).

This team's simulations are data-driven and help interpret observations from such space missions as Interstellar Boundary Explorer (IBEX), New Horizons, Ulysses, Voyager, and a fleet of near-Earth spacecraft.

Methods & Codes

These simulations model the flow of atoms with a few systems of the Euler gas dynamic equations describing the different atom populations dependent on the domains of their origin.

These are components of a Multi-Scale Fluid–Kinetic Simulation Suite (MS–FLUKSS)—an adaptive mesh refinement code we have built on the Chombo framework

Why Blue Waters

Results & Impact

- Performed a time-dependent simulation of the heliosphere that produces a comet-like heliotail and showed that the 11-year solar cycle leads to the formation of ENA lobes with properties remarkably similar to those observed by IBEX;

- Reproduced observations of pickup ions by New Horizons. These results have been published in six papers (one more paper is in press and three are in preparation) and reported at more than 10 scientific meetings

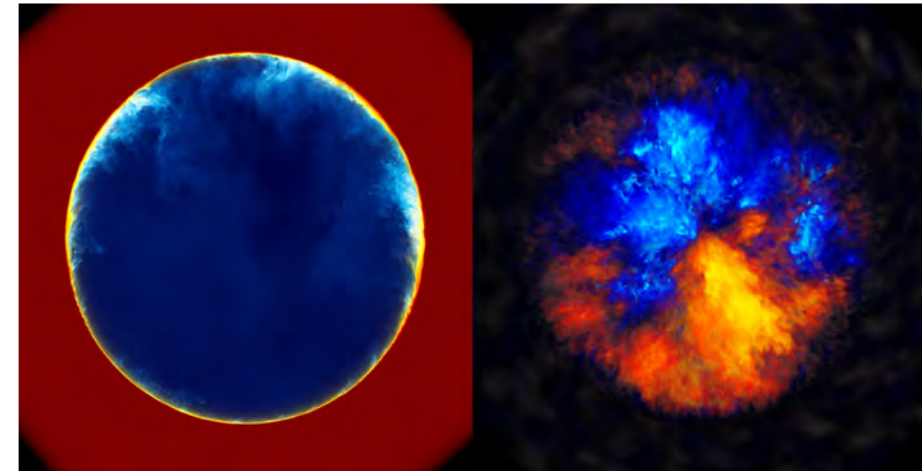- This research has been highlighted by the American Astronomical Society and other web news outlets.

The Blue Waters system and project staff responded in a timely manner to our concerns and were very helpful in the development of job-scheduling strategies. The overall performance and reliability of Blue Waters has been outstanding.

Allocation: NSF PRAC/4000 Knh
PI: Paul Woodward
University of Minnesota
Space Science

3D STELLAR HYDRODYNAMICS OF H-INGESTION FLASHES AND SHELL MERGERS

Research Challenge
The project is simulating brief events in the interiors of evolved stars that result in the ingestion of unprocessed new fuel into convection zones above nuclear-burning shells. The new fuel can burn very violently under the much hotter conditions in the convection zone after reaching a sufficient depth within it. This burning sets off a series of reactions that dramatically affects the nucleosynthesis of heavier elements and hence affects the ultimate expulsion of heavier elements into the surrounding interstellar gas.

Snapshots of the distribution of ingested gas from above the convection zone generated by core hydrogen burning in a model 25 solar mass star. We see a prominent dipole-like convection flow. Left, entrained gas. Right, radial component of velocity (outward is red and yellow, inward is blue and white). Flow goes directly through the center of the star.

Methods & Codes
The project used the Piecewise–Parabolic Method (PPM) for gas dynamics coupled with the Piecewise–Parabolic Boltzmann (PPB) moment-conserving advection scheme for the multifluid volume fraction.

Results & Impact
The project is producing a database of detailed simulations that investigates the phenomenon of convective boundary mixing at unprecedented accuracy for convection zones that extend over ranges in radius of more than a factor of two. Global convection modes play an important role in these situations, making simulation difficult and costly. Convective boundary mixing plays an important role in stellar evolution.

Why Blue Waters
Blue Waters provides a special ability to run at a sufficiently large scale that large computations can each be completed in less than one week. This allows the team to pose questions and get answers on a timescale that is conducive to productive thought and dynamically adjust research direction.
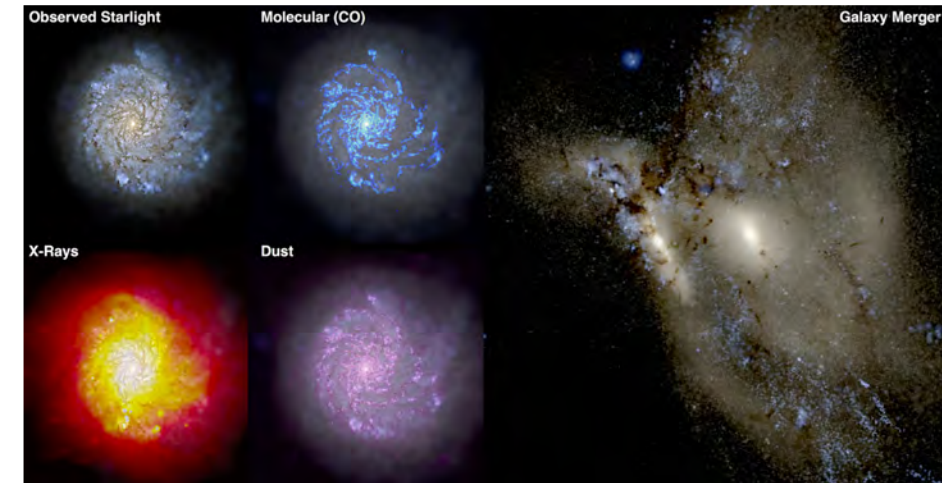
USING BLUE WATERS TO UNDERSTAND THE ORIGINS OF GALAXIES AND THE NATURE OF DARK MATTER

Research Challenge

At a fundamental level, the study of galaxies and stars seeks to answer the question, "How did we get from the Big Bang to the Milky Way?" This is an immensely challenging question involving the interplay among gravity, fluid dynamics, radiation and matter, and stars exploding as supernovae, giving rise to explosive outflows of material from galaxies that can reach across the observable universe.

The research team seeks to understand the origin and nature of galaxies, using massively parallel simulations that follow the birth and evolution of galaxies and stars from the very early universe to the present day.

Simulations of a Milky Way-like galaxy. Observed starlight (mock images) is presented with overlaid intensity maps showing mock carbon monoxide (molecular gas), X-rays, and dust emission. At the right is a mock Hubble image during a galaxy collision, where violent bursts of star formation are triggered.

Methods & Codes

The research team has run a large suite of cosmological, high-resolution simulations including detailed treatments of the physics of the interstellar medium, star formation, feedback in radiation and supernovae, magnetic fields, and cosmic rays. The simulations use the Feedback In Realistic Environments (FIRE) physics methods in the GIZMO code, a new massively parallel multimethod, hybrid Lagrangian–Eulerian finite-element, high-order, radiation-magnetohydrodynamics code (unique in numerical methods and physics-supported).

Results & Impact

Our cosmological simulations target galaxies from the faintest dwarfs through the Milky Way at the ultra-high-resolution and realism required to leverage the next generation of observations. These simulations model the physics of galaxy formation, uniquely incorporating not only all of the important stellar feedback mechanisms but also magnetic fields, physical (anisotropic) Braginskii conduction and viscosity, passive scalar (metal) diffusion, and explicit, multiwavelength radiation hydrodynamics. It has revealed fundamental new insights into how stars alter their galactic environments and has changed our observational inferences about the nature of dark matter in those galaxies.
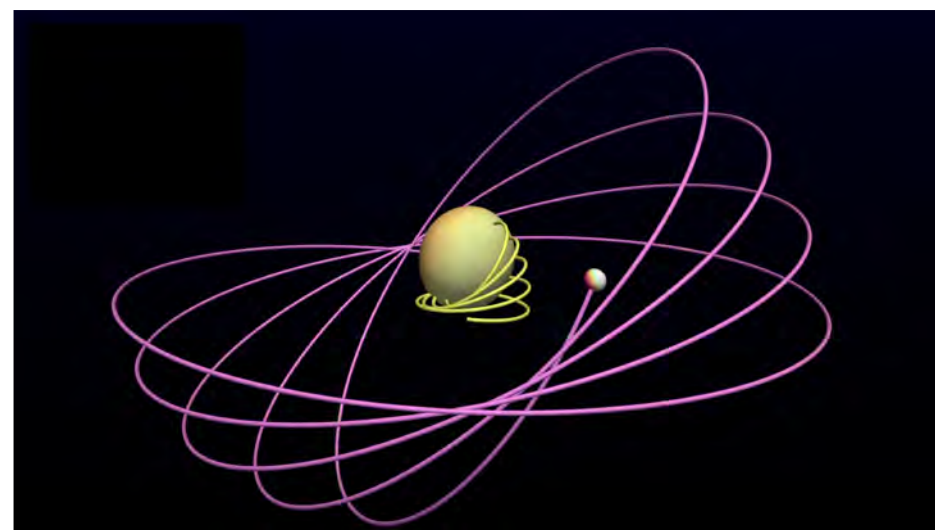
Why Blue Waters

Blue Waters is critical for this research because the enormous computational challenges enumerated above require >100 million CPU-hours on tens of thousands of processors and requiring tens of terabytes of active memory to store and evolve the immensely complex physical systems, and the simulations produce petabytes of data products. No other current facility enables this research.

Allocation: NSF PRAC/8200 Knh
PI: Saul Teukolsky
Cornell University
Space Science

PETASCALE SIMULATIONS OF MERGING BLACK HOLES AND NEUTRON STARS

Two spinning black holes in orbit around each other, calculated by full solution of Einstein's equations of general relativity. In Newtonian gravity, the orbits would stay in a plane. Because the holes are spinning, the orbits precess (tilt).

Research Challenge

The primary scientific objective of this project is to theoretically underpin and improve the ability of LIGO (Laser Interferometer Gravitational-Wave Observatory) to extract the rich information that the observed GWs (Gravitational waves) carry. GWs provide a new window on the universe that will enable us to test current understanding of fundamental physics as well as learn about the most extreme events in the cosmos.

Methods & Codes

- Most of the computations are done with the SpEC code (Spectral Einstein Code) for treating black holes.
- Also developing a new code, SpECTRE, with innovative methods to treat neutron star systems.

Results & Impact

This team will release a new version of public catalog of gravitational waveforms for use by all researchers, largely through simulations on Blue Waters. The new version will increase the size of the catalog from 174 waveforms to well over a thousand. These waveforms have already been used to produce a very accurate waveform model that LIGO can use in its data analysis.

Why Blue Waters

The numerical code runs most efficiently on 50 to 70 processors for each waveform. Blue Waters' nodes are perfectly sized for the project to use one or two nodes per waveform and explore hundreds of different parameter values to develop our catalog.
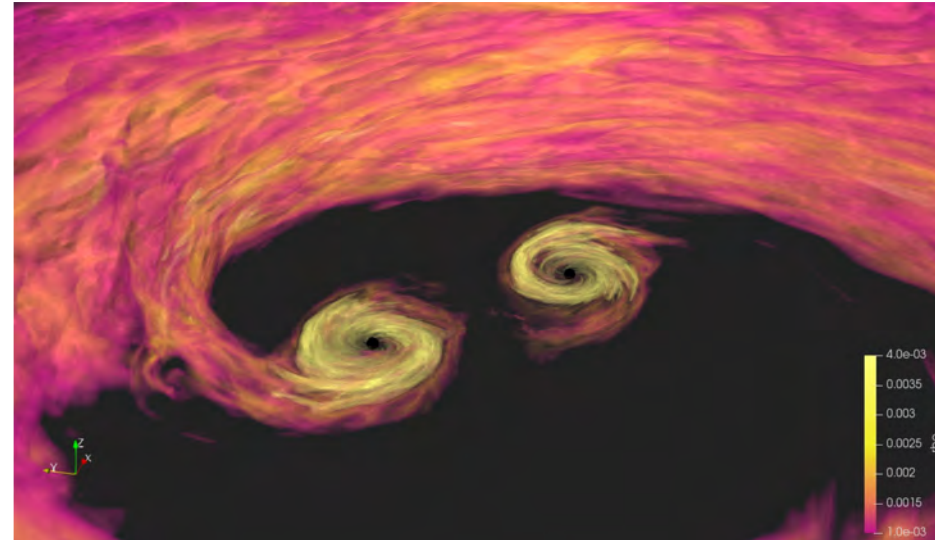
Magnetohydrodynamic simulation of mini-disk and circumbinary disk interactions

SHEDDING LIGHT ON INSPIRALING BINARY BLACK HOLES WITH MAGNETIZED MINI-DISKS

Research Challenge

Realistic accretion disk simulations are particularly challenging as they involve a multitude of physical processes interacting over large dynamic ranges in space and time. In actual systems, gas is collected at scales a million times larger than the black holes themselves and yet many cells per black hole width must be used to capture the relativistic plasma dynamics in their vicinity. Consistency between the gas's thermodynamics and radiation model is desirable to produce self-consistent predictions of the light produced by the modeled systems, which is the ultimate goal of our program

Methods & Codes

- A flux-conservative, high- resolution, shock-capturing general relativistic magnetohydrodynamic (GRMHD) code, HARM3d, is used.

- HARM3d is written in a way so that any metric or coordinate system may be adopted, which accommodated the implementation of a novel, time-dependent, non-uniform gridding scheme to resolve the huge scale differences for features near and far away from black holes.

Results & Impact

- The inclusion of magnetic stresses provides greater realism to how gas is brought in from large distances and how it dissipates its orbital energy.

- A new phenomenon is discovered where the irregular circumbinary flow can modulate the rate of accretion onto the mini-disks .

- A first-of-a-kind radiative transfer calculation in time-dependent general relativity using the simulation's data as an emitting source, has resulted in the first electromagnetic spectrum of accreting supermassive black holes in the inspiral regime.

Why Blue Waters

Blue Waters' capability and support has made the 3D GRMHD mini-disk simulation, which ran for three orbital periods and used 18 million floating-point-core-hours or 1.2 million node-hours, possible. The simulations used 600 × 160 × 640 or approximately 60 million cells with about 3 million timesteps using 600 nodes or 19,200 Blue Waters cores. We have further benefited from their visualization team's efforts in producing state-of-the-art visualizations of our simulations.
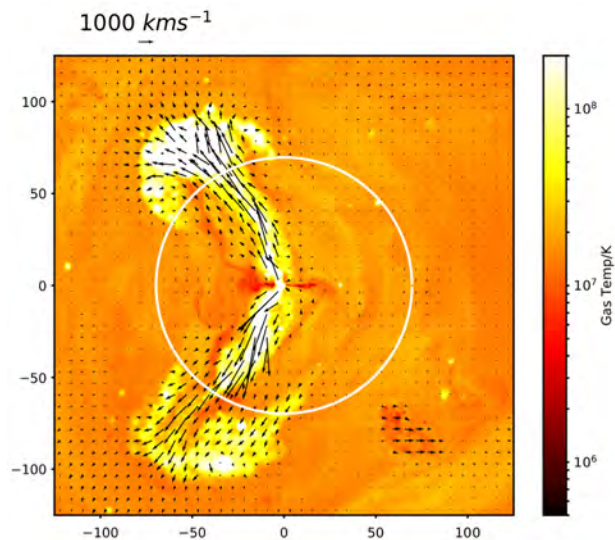
Outflows from the active galactic nucleus (AGN) of the central cluster galaxy. The columnation of the flow is emergent from the simulation: the energy from the AGN is injected isotropically. This outflow counteracts the cooling of the intracluster gas and suppresses star formation.

UNIFIED MODELING OF GALAXY POPULATIONS IN CLUSTERS

Research Challenge

Understanding the physical processes that occur in group and cluster galaxy environments is key to gaining insights into the evolution of baryons and galaxies across the age of the universe. Clusters will provide very tight constraints on the understanding of galactic feedback processes. They are also key probes of cosmology and large-scale structure. However, using clusters as cosmological probes requires understanding of the relationship between observables and the total mass of the cluster, which in turn requires the detailed modeling of the gravitational/hydrodynamic processes using large simulations.

Methods & Codes

This project uses the highly scalable N-body/hydrodynamics code ChaNGa to model the formation and evolution of a population of galaxies in a Coma-sized galaxy cluster. ChaNGa is built on the Charm++ infrastructure, and leverages the object-based virtualization and data-driven computation style of Charm++. ChaNGa has been shown to scale well to one-half million cores on Blue Waters. This project's simulations will be compared to observations of cluster galaxies to understand the physical and temporal origin of their morphologies.

Results & Impact

This project has completed simulations of several smaller galaxy clusters, in preparation for their flagship simulation. Even the completed smaller simulations are advancing the state of the art in the simulation of galaxy clusters. Models based on energy injected from supernovae and active galactic nuclei in field galaxies are able to naturally explain the properties of clusters and the galaxies within them. In this way, the models become predictive of the growth of galaxies and the black holes that power the active galactic nuclei.
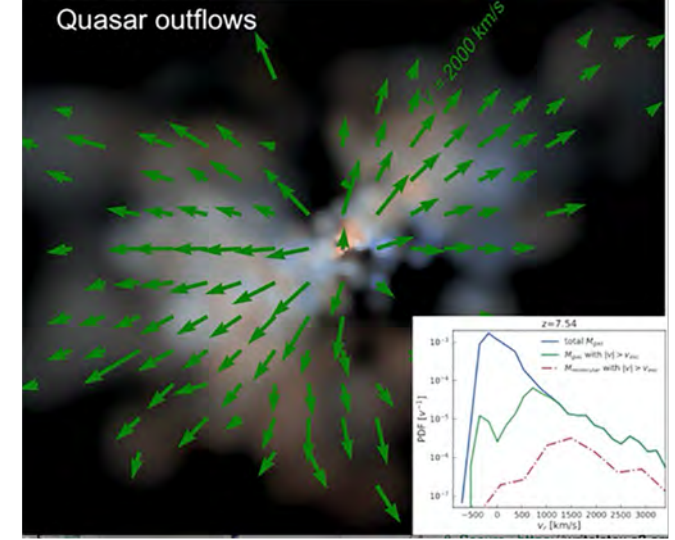
Why Blue Waters

The project's scientific goals require modeling over a large dynamic range in mass and space. It has been demonstrated that mass resolutions of order 10^5 solar masses are needed to accurately follow star formation and galaxy morphology. Likewise, it is necessary to model a galaxy cluster of order 10^{15} solar masses that is comparable to those observed over a range of redshifts. Hence, 10 billion particles are needed. Such a simulation can only be run on the largest computers available. Furthermore, the long-range nature of gravity requires a high-performance, low-latency network to perform the calculation.

Allocation: NSF PRAC/1856 Knh

PI: Tiziana DiMatteo

Carnegie Melon University

Space Science

Conical gas outflows around the most massive black hole are powered by the quasar, which energizes and blows gas out from the host galaxy and also quenches star formation. The insert shows a histogram of the gas velocities around the black hole. The simulations predict this outflow should be observable around the first massive black holes

TINY GALAXIES HOST THE FIRST GIANT BLACK HOLES: BLUE TIDES SIMULATION MAKES CONTACT WITH THE FIRST 700 MILLION YEARS OF COSMIC HISTORY

Research Challenge

Understanding and detecting the first galaxies and black holes formed in the first billion years, is one of the main observational and theoretical challenges in galaxy formation. The fundamental challenge in simulations trying to understand this epoch of the Universe is that extremely large volumes need to be analyzed as the first objects are rare, while at the same time extremely high resolution is required as the first galaxies and quasars are expected to be small and compact. A complete simulation of the universe at the epochs being studied, requires a small enough particle mass to model the first galaxies. It also requires an enormous volume, of the order of 1 cubic gigaparsec (1 Gpc^3 is 3×10^{28} cubic light years) to capture the rarest and brightest objects. The first requirement is therefore equivalent to a high particle density, and the second to a large volume. The BlueTides (BT) simulations on Blue Waters have successfully answered these challenges.

Methods & Codes

- The BT simulations have been made possible with the use of the new cosmological hydrodynamic simulation code (MP)-Gadget which is massively parallel
- Radical updates to the code efficiency, the Smooth Particle Hydrodynamics formulation, and star formation modeling, have met the challenge of simulating the next-generation space telescope fields and the effective use of the full BW machine.

Results & Impact

- BT is the only simulation making contact with the recently discovered supermassive black hole and can now make predictions regarding its formation, history, and observation signatures by next-generation telescopes. BT is also the first and only cosmological simulation of structure formation that has run on the full BW machine.
- The host galaxy of the most distant known quasar (from 690 million years after the Big Bang), was explored in the large-volume cosmological hydrodynamic simulation.
- BT simulations studied the feedback around the highest redshift quasar in and predicted that there are significant outflows observable at radio wavelenghs around this quasar.
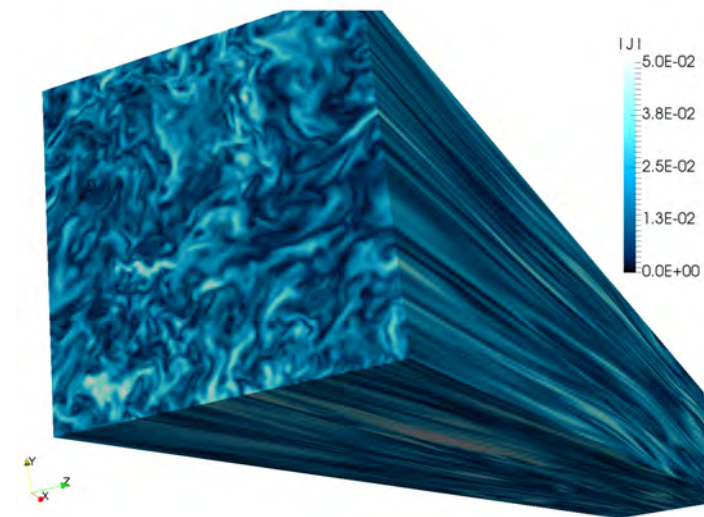
Why Blue Waters

- The size of the simulations being performed, meant that BW was the only system that could solve it, both in terms of memory and compute cores. Simulations essentially used the whole BW.
- Assistance by BW team in scheduling, file system tuning and I/O, and run time problem analysis, were also very important and beneficial.

KINETIC SIMULATIONS OF PLASMA TURBULENCE

Visualization of small-scale structure in kinetic plasma turbulence. The simulation was conducted using a newly developed version of spectral, fully implicit code SPS.

Research Challenge

Plasma turbulence plays a significant role in the dynamics of many systems in the universe, from laboratory fusion experiments to the Sun and astrophysical objects such as accretion disks. While parameters, geometry, and some aspects of the physics may differ among these systems, there is also a large degree of universality, which makes understanding plasma turbulence a grand challenge problem relevant to many fields of study.

Methods & Codes

Due to the multiscale nature of plasma turbulence (different processes dominate at different scales), the team runs simulations that investigate plasma on a microscopic scale. Two codes, providing complementary approaches to describing plasmas via the Vlasov–Maxwell equations, are used:

- VPIC – a general-purpose particle-in-cell (PIC) plasma simulation code
- SpectralPlasmaSolver (SPS) – a fully implicit code based on a spectral decomposition approach that possesses exact conservation laws for long-term, accurate simulations

Why Blue Waters

The simulations needed to address the scientific questions of this project solve a global problem that cannot be split into a series of smaller simulations. As such, they require large memory, fast on-node computation, and fast internode communications. For this reason, they require a high-performance computing resource like Blue Waters and cannot be conducted on cloud resources.

Results & Impact

Most prior investigations of plasma turbulence have focused on plasma regimes similar to those found in the solar wind at the Earth's orbital distance from the sun. But with the summer 2018 launch of the Parker Solar Probe (PSP) – a "NASA mission to touch the Sun" – there is now an urgent need to develop models for plasma turbulence in the regimes near the sun that will be encountered by PSP. Through 3D SPS and 2D VPIC simulations, the team has confirmed the existence of this type of theoretical new regime of plasma turbulence (simulation results match will with theoretical predictions) that should be characteristic of plasma found close to the sun.

Allocation: NSF PRAC/1800 Knh

PI: Yi Hsin Liu

Dartmouth College

Space Science

3D NATURE OF COLLISIONLESS MAGNETIC RECONNECTION AT EARTH MAGNETOPAUSE

In (a), the total current density on a 2D plane where y = 0. In (b), the x–y cut of the current density across the location of the intense current near the x-line. Similarly, in (c), the reconnected field; in (d), the electron outflow; in (e), the ion outflow.
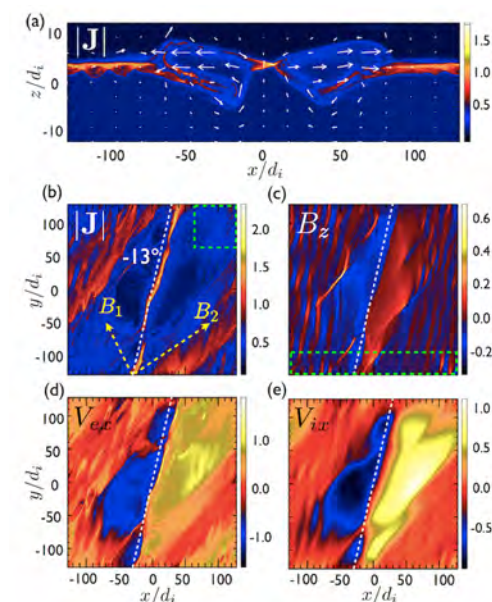
Research Challenge

The overall goal of this project is to develop an adequate understanding of the 3D nature of asymmetric magnetic reconnection, which will help to advance the forecasting of space weather. To this end, the team aims to answer the question: Is there a simple principle that determines the orientation of the reconnection x-line (the line along which magnetic field lines reconnect) in such an asymmetric current sheet? The solution to this problem remains unclear according to the current understanding of magnetic reconnection.

Methods & Codes

The particle-in-cell code VPIC is used for magnetic reconnection simulations, and the open-source package ParaView is used for visualization of the results. VPIC solves the relativistic Vlasov–Maxwell system of equations using an explicit charge-conserving approach. Charged particles are advanced, and then the current and charge density are accumulated on grid points to update electromagnetic fields.

Results & Impact

The team studied the orientation and stability of the reconnection x-line in asymmetric geometry in 3D systems. Some simulations were used to determine the x-line orientation that maximizes the reconnection rate. Another experiment suggested that reconnection tends to radiate secondary oblique tearing modes if it is externally (globally) forced to proceed along an orientation not favored by the local physics, which leads to turbulence inside small periodic systems. These and other simulation results have been used to interpret observations by the Magnetospheric Multiscale Mission (MMS), and could also be used in conjunction with the upcoming Solar wind Magnetosphere Ionosphere Link Explorer (SMILE) mission.

Why Blue Waters

Because the x-line has a dimension down to electron scale, a fully kinetic description is necessary. Given the available computational capability, it has become possible to use a first-principle kinetic simulation to investigate the dynamics of the x-line in a reasonably large 3D system, which spans from electron kinetic (micro) scale to magnetohydrodynamics (macro) scale. A representative 3D run in this project simulates 2 trillion charged particles on 6 billion grids, generating hundreds of TBs of data for each run. Blue Waters not only provides the computational resource for the calculation but also the online storage for the output and restart files.
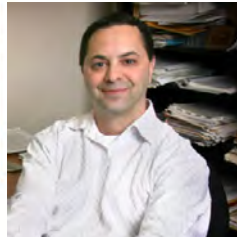
Geoscience

Allocation: BW Professor/210 Knh
PI: Larry Di Girolamo
University of Illinois at Urbana-Champaign
Geoscience

THE TERRA DATA FUSION PROJECT
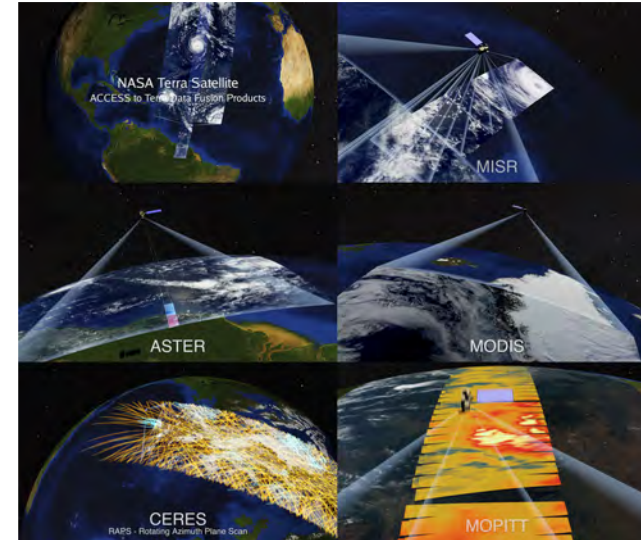
Images from an animation that dynamically displays and projects the radiance imageries generated from a single Terra basic fusion granule (Orbit 3671) onto Earth as observed from all 5 by Terra instruments. Video at https://youtu.be/C2uyjRGwwOs.

Research Challenge

This work benefits users of data from NASA's Terra instrument through the fusion of data from the five Terra instruments. The need for data from Terra to accelerate analytical applications that serve the scientific community, governmental and commercial needs, and the educational community has never been greater. The goals of the Terra fusion products are to facilitate ease of use and accuracy through data fusion, reduce errors and redundancy for users of data from Terra, and to provide a framework for data fusion that could extend to other NASA projects.

Methods & Codes

- Transfer the entire Terra record (Level 1B radiance; >1 petabyte) to Blue Waters from NASA centers
- Develop software for whole-mission processing to create fusion products
- Optimize data granularity and HDF packaging for parallel I/O
- Distrubute Terra fusion products through existing NASA services

Results & Impact

- Successfuly transferred terra data from 5 NASA sites and produced 2.4 PB of Terra full-mission fused data products
- Carried out science investigations using the Terra fusion data set resulting in radically different understanding of the distributions of cloud drop sizes in our atmosphere than those previously available, and that have been verified by spot measurements in field campaigns
- Developed a tool too dynamically visualize Terra data onto 3D earth

Why Blue Waters

The large storage, IO bandwidth, and computing facilities of BlueWaters provides an optimum framework for large-scale processing, analytics, and mining of the entire Terra record. In addition, the BlueWaters project staff provides critically needed expertise to optimize the Terra fusion workflows.

Allocation: BW Professor/240 Knh

PI: Kaiyu Guan

University of Illinois at Urbana-Champaign

Geoscience

FORECASTING GLOBAL CORP PRODUCTIVITY THROUGH INTEGRATING NOVEL SATELLITE DATA AND PROCESS-BASED MODELS

Conceptual frameworks for STAIR fusion algorithm.
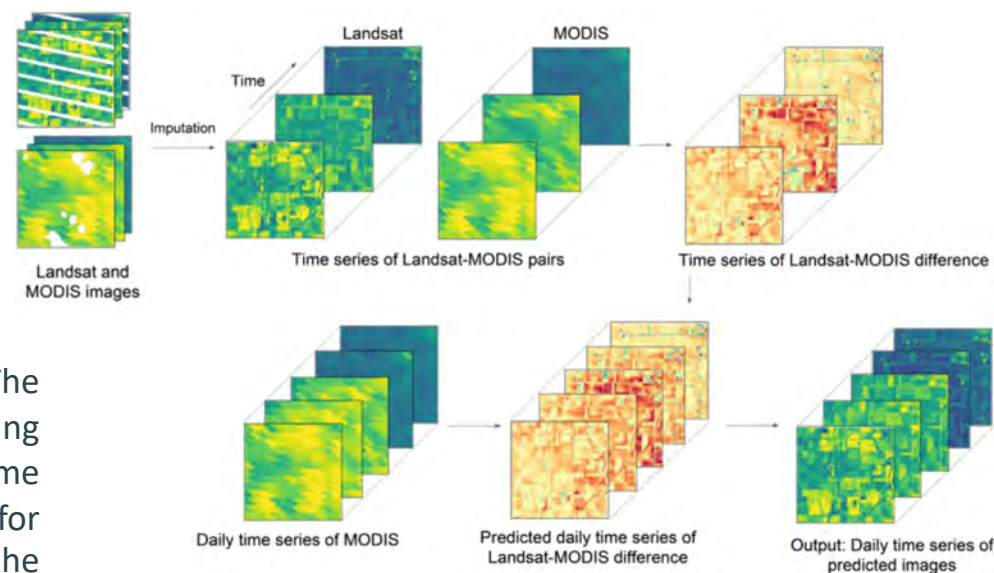
Research Challenge

Global food security is under continuing pressure from increased population and climate change. The goal of this project is to improve the predictability of regional and global crop yields by integrating advanced remote-sensing and process-based modeling. Satellite observations can provide real-time monitoring of crop acreage and growth condition. Process-based crop models are effective tools for studying the impacts of climate change on regional and global food production, for assessing the effectiveness of adaptations and their potential feedback to climate, and for attributing different pathways through which climate can impact crop yields.

Methods & Codes

We developed new multi-satellite (Landsat, Sentinel, MODIS, VIIRS) data fusion algorithm STAIR to generate high spatio-temporal resolution (30m and daily) and gap/cloud free satellite images for the contiguous United States. We trained machine learning algorithms for within-season crop type classification based on time series of satellite data. We improved the crop growth representation in the Community Earth System Model (CESM), and used SALib package for parameter sensitivity analysis and PyDREAM package for Bayesian parameter calibration.

Why Blue Waters

Blue Waters is essential for our research since other resources, such as those available from XSEDE, are not suitable for our project considering the petabyte-level storage demand, data availability, and intensive I/O and computational demands.

Results & Impact

Our results show that STAIR can provide good quality fused data for crop monitoring. By using time series of satellite data, our machine learning approach showed a relatively high overall accuracy of 96% for classifying corn and soybeans from 2000 to 2015, and achieved 95% overall accuracy for classifying corn and soybeans by late July of each year. The new process-based model (CLM-AgSys) corrected deficiency of CLM in simulating carbon allocation and showed better performance in simulating yield as well as surface fluxes.

EXTREME CONVECTIVE STORMS UNDER CLIMATE CHANGE

Research Challenge

The future impact of thunderstorms and thunderstorm systems such as hurricanes under a globally warmed climate are still uncertain. PGW (pseudo global warming) simulations show that hurricanes under climate change appear to be more hazardous in terms of inland flooding and tornado generation. This research is important from the perspective of basic science but will also help to inform decision-makers such as emergency managers on how to prepare for future disasters.

Methods & Codes

Used an event-based implementation of the PGW methodology. Modified atmospheric states drawn from GCM (global climate model) output were applied to constrain WRF (Weather Research and Forecasting) model simulations at high resolution. Supplemented these PGW simulations with idealized simulations using CM1 (Cloud–Model 1).

Both WRF and CM1 are community codes.

Results & Impact

PGW simulations exhibited more intense TCs (tropical cyclones). The TCs under PGW also produced significantly more accumulated rainfall. These and other experiments lend support to the hypothesis that an increase in sea surface temperature due to human-induced climate change will intensify landfalling TC , which in turn will result in more numerous tornadoes.

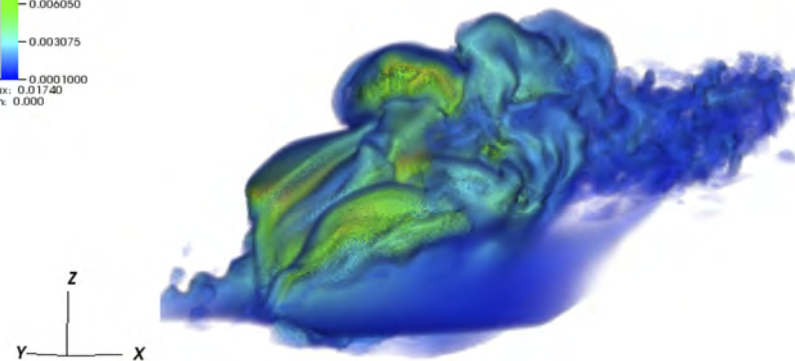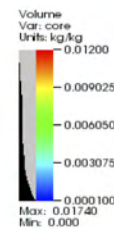## Why Blue Waters

This research requires very large geospatial domains that have fine grid point spacings and long-time integrations with high rates of model output. Moreover, quantifications of uncertainty require that such realizations be repeated over multiple experiments. The Blue Waters allocation is providing us with the resources needed to achieve this unprecedented level of climate simulation.

Allocation: BW Professor/275 Knh
PI: Sonia Lasher-Trapp
University of Illinois at Urbana-Champaign
Geoscience

"BREATHING" CLOUDS AND STORMS:
INFLOW AND ENTRAINMENT, OUTFLOW AND PRECIPITATION

Research Challenge

"Entrainment" is the process by which the turbulent motions within clouds bring dry air from outside the cloud inward. In time, entrainment can limit storm development and precipitation, but is not always effective in doing so. Long-standing problems in meteorological models have been to understand why they tend to predict rain formation too early, and/or in excessive amounts, and why models often miss predicting outbreaks of storms. This project investigates if these deficiencies in past models are related to poor representation of entrainment, and its influence upon precipitation production of storms.

Precipitation mass in a thunderstorm developing in an environment where the wind speed increases greatly with height. Much of the mass from the older part of the cloud (blue colors on right side) was depleted by entrainment.

Methods & Codes

- NCAR's CM1 model was used to simulate convective clouds and storms at high resolution.
- The NSSL microphysical scheme within CM1 was used to model the details of precipitation formation.
- Entrainment was evaluated with a project-developed code that calculates mass fluxes into the core of the storm as it evolves.

Results & Impact

The latest results from this project suggest that:

- Developing thunderstorms growing in a scenario where the winds increase strongly with height, initially, entrain three to five times more dry air.
- Storms developing closer together may precipitate less initially, but if they do precipitate, their outflow can force new storms that precipitate much more.
- The amount of large ice particles that fall from storms appears to be most important for strengthening their outflows, rather than the evaporation of rain, as is sometimes assumed.

Why Blue Waters

Blue Waters has been essential for achieving the high resolution required within a given simulation to properly represent the smaller cloud motions that are important for entrainment and precipitation development over the larger spatial and temporal domains required for thunderstorms and their outflows. As a result of its huge number of nodes, its high speed, large memory, and its large storage capability for high-resolution model output and analysis, Blue Waters enables the team to conduct detailed calculations over millions of grid points. The hardware needed to run these simulations quickly supersedes the limits of most computers.

Allocation: GLCPC/540 Knh

PI: Jennifer Corcoran

University of Minnesota, Twin Cities

Geoscience

Top left: June 2016 WorldView-3 image; Top center: Stereo derived Digital Surface Model (DSM) from the image date on the left; Bottom left: August 2016 WorldView-3; Bottom center: Stereo derived DSM from the image date on the left; Right: Difference raster between the two DSMs in the center.

IMAGE PROCESSING TO BUILD A MULTI-TEMPORAL VEGETATION ELEVATION ECOSYSTEM MODEL (MTVEEM) OF THE GREAT LAKES BASIN (GLB)
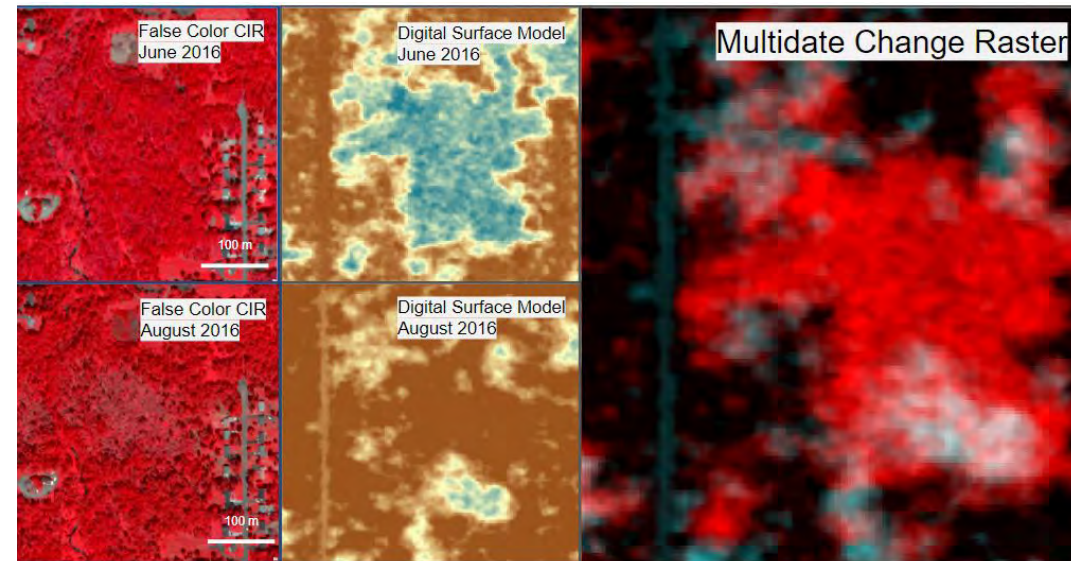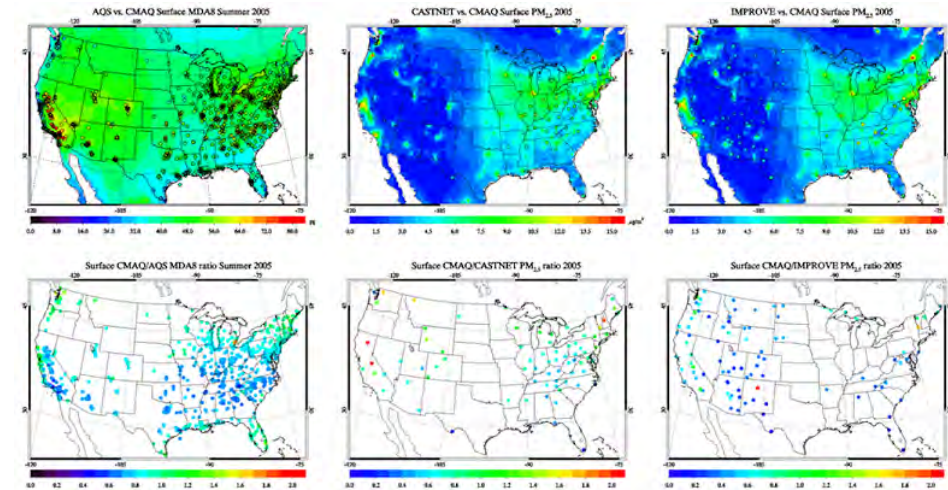
Research Challenge

How are the ecosystems across the GLB changing? Stereo submeter satellite imagery is used for the generation of multi-temporal, highly accurate surface elevation models which are highly demanded by natural resource managers. This research will identify both natural changes (e.g. flooding, forest blowdown, fire, insect and disease damage) and anthropogenic changes (e.g. harvest and land-cover change). MTVEEM will improve habitat and biological modeling for mapping and monitoring canopy change in forested and wetland habitats across the GLB.

Methods & Codes

*In 2017, 83,000 stereo pairs were processed; in 2018 50,000 pairs are expected

*Each job converts the input image pairs to GeoTIFF and calls the elevation extraction software (SETSM)

*Each pair is run on a single node, submitted in batches of 2-100 tasks; each task takes an average of 12 NH (6 NH charged due to low-priority discount)

*Additional 150,000–200,000 NH to process and classify ortho-images

Why Blue Waters

The amount of stereo imagery the GLB and the computational burden to process this stereo imagery is well beyond those available from academic, private, and government systems. Since these image sets are flexible for processing, these researchers work with the Polar Geospatial Center to backfill using Blue Waters existing and new stereo imagery over the entire GLB. Each stereo pair is about 30GB and the total number of pairs processed to date is about 100,000 which moves this volume alone into the petascale environment.

Results & Impact

The final product of this research, a seamless and registered surface elevation ecosystem model (MTVEEM) across the GLB, will enable a large range of science activities and provide substantially higher resolution data than currently available. These canopy maps and change detection products will provide positional accuracies of less than a couple meters with the added ground control points. Preliminary results show great promise for providing valuable data to myriad of coastal and terrestrial ecosystem science researchers and decision-makers across the entire GLB.

Allocation: Illinois/550 Knh
PI: Donald Wuebbles
University of Illinois at Urbana-Champaign
Geoscience

PARTICULATE MATTER PREDICTION AND SOURCE ATTRIBUTION FOR U.S. AIR QUALITY MANAGEMENT IN A CHANGING WORLD

Research Challenge

The research examines global changes in climate and emissions, primary particulate matter and ozone. The study's objectives are to better understand how global changes impact U.S. air quality, to project future trends, and to quantify pollution sources and assign their attributions: natural vs. human-influenced, national vs. international agents, natural variations vs. climate change. Researchers can then provide actionable data for U.S. environmental planners and decision makers for use when designing management strategies, including local controls, domestic regulations, and international policies. A state-of-the-science prediction system will be used for three primary experiments: historical simulations to validate the system, future projections to quantify impact of changes in global climate and emissions, and sensitivity analyses to determine future changes in pollution sources and their relative contributions each sources' attributions.

Comparison of EPA Air Quality Standards observations and Community Multiscale Air Quality (CMAQ) simulations. Ozone values are from summer JJA MDA8 and PM2.5 measurements are the annual mean. Dots shows AQS/CASTNET/IMPROVE observations and background shows CMAQ results (upper panel). Lower panel shows the ratio of observations to CMAQ results.

Methods & Codes

The dynamic system couples a global climate–chemical transport model with regional climate and air-quality models over North America, to determine individual and combined impacts of global climate and emissions changes on U.S. air quality. These include uncertainty evaluations, from the present to 2050, under multiple climate and emission scenarios. The results from the global and regional model simulations for the past are evaluated with observational data to assess the capabilities of the model simulation and impacts of emissions change, climate change, and long-range transport on future U.S. air pollution.

Results & Impact

The system's models were used to simulate major air pollutants from 1980 to 2005. Evaluation of the system using historic EPA measurements showed that it can capture the distribution of ozone pollution with substantial underestimations in urban and suburban areas, and the pollution pattern of particulate matter, while some isolated sites had substantial discrepancies. These results show the capability of the system. Further improvements planned by the researchers will improve the predictive capability of the system, and produce a more complete scientific understanding of global climate and emissions changes imposed on U.S. air quality management and a more reliable projection of future pollution sources and attribution changes.

Why Blue Waters

The computational demand of high-resolution climate models used in this project is very extensive. In addition, we are using a fully coupled model of the Earth's climate system with interactive chemistry, which is also computationally expensive even when not run at high resolution. Blue Waters, with its petascale computational facility, large number of nodes, and storage capability for the output from the high-resolution model simulation, is essential for our project. Blue Waters has given us the computational resource, data management, and support staff to perform our research.

Allocation: Illinois/240 Knh

PI: Matthew West

University of Illinois at Urbana-Champaign

Geoscience
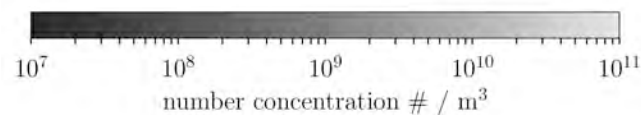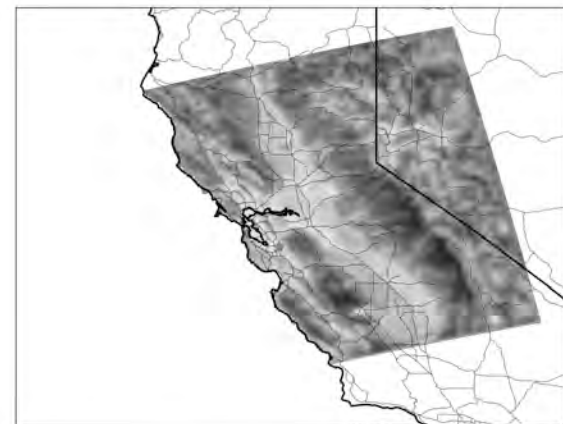
Figure: Horizontal distribution of particle number concentration located near the surface after 12 hours of simulation.

Simulating Aerosol Impacts on Climate, One Particle at a Time: A Regional-Scale, Particle-Resolved Aerosol Model to Quantify and Reduce Uncertainties in Aerosol–Atmosphere Interactions

Research Challenge

This research aims to reduce key uncertainties in quantifying the impact of atmospheric aerosol particles on Earth's climate. Aerosols profoundly impact the large-scale dynamics of the atmosphere because they interact with solar radiation—both by scattering and absorbing light, and by forming clouds. Due to computational constraints, current models do not resolve individual particles and their microscale interactions so methods representing the high-dimensional and multiscale nature of aerosol populations apply large simplifications which also introduces unknown errors.

Methods & Codes

To overcome the current limitations in representing aerosols and associated uncertainties, two models were combined.

- **PartMC–MOSAIC -** a detailed aerosol model that tracks the size and complex composition of individual particles in the atmosphere.
- **3D regional Weather Research and Forecast (WRF)** - an advanced numerical weather model that captures the transport of chemical species in the atmosphere.

Results & Impact

WRF–PartMC is the only model of its kind, and this work is changing the field of aerosol science because it provides the first benchmark for more approximate models and captures the complex aerosol composition that current-generation models are unable to simulate. Results obtained from a particle-resolved aerosol simulation for a realistic, spatially resolved three-dimensional domain in California - 50 billion computational particles, including their compositional changes due to gas-to-particle conversion, their coagulation events, and their transport by wind and turbulence.

Why Blue Waters

Access to Blue Waters allows for a cutting-edge model formulation that pushes both science and computing by combining the large-scale features of state-of-the-art 3D models with the process-level physical representation of box models. Modeling 3D domains with as many as 100 billion tracked particles creates challenges due to computationally intensive equations per particle and memory requirements to track high-dimensional particle composition. To enable simulations of aerosols at both a high spatial and compositional resolution requires tens of thousands of cores, fast interconnections between those cores, and sufficient memory.

Allocation: Illinois/500 Knh

PI: Ryan Sriver

University of Illinois at Urbana-Champaign

Geoscience

ANALYZING TROPICAL CYCLONE-CLIMATE INTERACTIONS USING THE HIGH RESOLUTION COMMUNITY EARTH SYSTEM MODEL
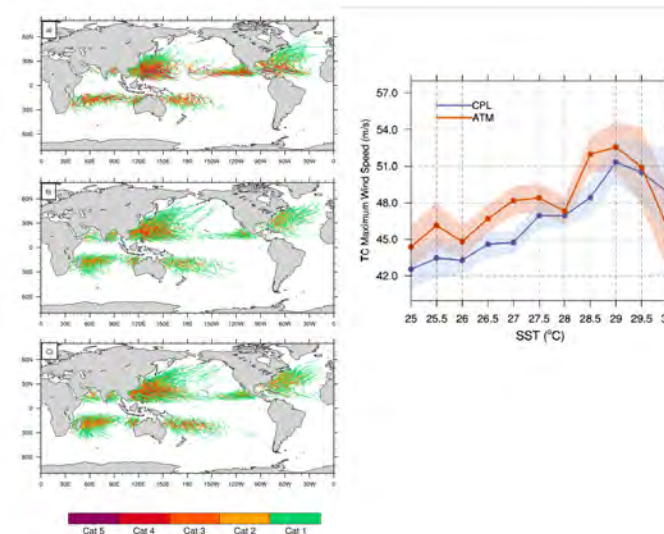
Research Challenge

Research has shown that tropical cyclones (TCs) are both influenced by and have a significant impact on the Earth's climate, particularly through their interactions with the upper ocean. Understanding these interactions is of great importance to obtaining a better knowledge of the Earth's climate system and constraining uncertainty in climate projections. To this end, the team sought to characterize their climate model's self-generated TCs, analyze the sensitivity of these simulated TCs to ocean coupling, and to diagnose the impact of the model's self-generated TCs on the global ocean within the modeled climate.

(Left) Thirty years of accumulated tropical cyclone (TC) tracks in the (a) observational best track during 1985–2014, (b) the coupled CESM simulation, and (c) the atmosphere-only simulation. (Right) Relationship between SST (sea surface temperature) and maximum TC wind speed in the fully coupled simulation (blue) and the atmosphere-only simulation (salmon).

Methods & Codes

In this project, the team analyzed the interactions between TCs and climate using a high-resolution configuration of the CESM (Community Earth System Model). First, they performed a 30-year fully coupled high-resolution simulation for use as a control. They then ran an atmosphere-only simulation to examine the sensitivity of modeled TCs to ocean coupling. And finally they investigated the impact of TCs on the global ocean using a set of ocean-only simulations with and without TC forcing.

Results & Impact

Results from the atmosphere-only simulation revealed that ocean coupling is essential to capturing realistic TC intensity and intensification, influencing the simulated annual TC number, spatial distribution, and storm intensity. Results from the ocean-only simulations revealed that TCs can influence global ocean temperature patterns, ocean energetics, and ocean heat transport. They also observed that TCs can strengthen certain ocean circulations, but the influence of such circulation changes appeared to be fairly small on a global scale.

Why Blue Waters

TC–climate research falls at the interface between weather and climate modeling, requiring high-resolution grid spacing to resolve weather-scale TC features, as well as global-scale coverage and decades of integration time. Blue Waters has the unique capability to handle the computational demand associated with running the model at ultra-high resolutions, including scalability to over 15,000 cores, high frequency input and output, and post-processing and visualization of model results.

Allocation: Illinois/500 Knh
PI: Ryan Sriver
University of Illinois at Urbana-Champaign
Geoscience

THE RESPONSE OF TROPICAL CYCLONE ACTIVITY TO GLOBAL WARMING IN THE COMMUNITY EARTH SYSTEM MODEL
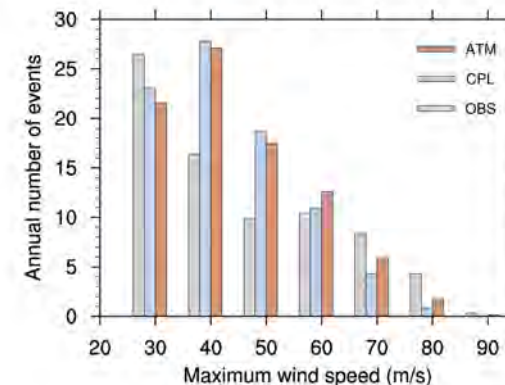
Research Challenge

Tropical cyclones are rare weather events, yet they consistently rank among the world's deadliest and costliest natural hazards. These cyclone–climate connections are poorly understood and largely missing from today's generation of Earth system models, yet they may be fundamentally important to understanding the mechanisms influencing climate variability and to improving projections of future climate change.
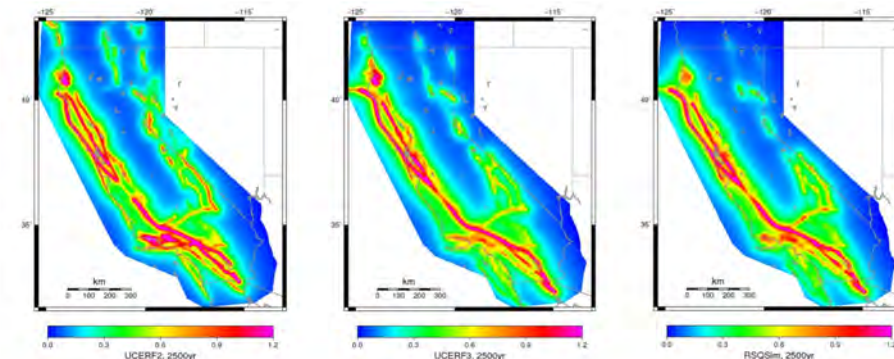

Binned frequency distribution of global annual average number of tropical cyclones for different maximum wind speeds from observations (grey), coupled CESM (steel blue), and atmosphere-only CESM (salmon)

Methods & Codes

The Community Earth System Model (CESM) is a comprehensive global climate model that consists of atmosphere, land, ocean, and sea ice components that are connected via a central coupler that exchanges state information and fluxes between the components. The Community Atmosphere Model has improved microphysics and cloud properties, and the prognostic modal aerosol package is activated in the current experiments.

Results & Impact

In this research project, the team is building on their recent work analyzing the relationship between TCs and climate using a high-resolution, state-of-the-art Earth system model. The comprehensive suite of model experiments have provided them with thousands of simulated cyclone tracks for different climate conditions and coupling configurations using a dynamically consistent Earth system modeling framework, enabling robust assessment of cyclone activity and variability in response to changes in climate.

Why Blue Waters

Tropical cyclones (TC) climate research falls at the interface between weather and climate modeling, requiring high-resolution grid spacing to resolve weather-scale TC features, as well as global-scale coverage and decades of integration time. Blue Waters provides the unique capabilities of handling the computational demand associated with running the model at ultra-high resolutions, including scalability to over 15,000 cores, high frequency input and output, and postprocessing and visualization of model results.

Allocation: NSF PRAC/9620 Knh

PI: John Vidale

University of Southern California

Geoscience

These maps show estimated shaking hazard produced using the RSQSim earthquake catalog as compared to the UCERF2 and UCERF3 hazard models. Maps show Peak Ground Acceleration with a probability of exceedance of 2% in 50 years. Units are in fractions of the acceleration of gravity, g. (a) UCERF2, (b) UCERF3, (c) RSQSim model.

SIMULATING ALTERNATIVE CALIFORNIA EARTHQUAKE HISTORIES USING RATE AND STATE FRICTION

Research Challenge

The Southern California Earthquake Center (SCEC) conducts an earthquake system science research program developing physics-based, predictive models of earthquake processes to improve probabilistic seismic hazards assessments. SCEC researchers use physics-based computational models, observation-based 3D earth structure models, and high-performance computer resources to improve probabilistic seismic hazard forecasts for California.

Methods & Codes

SCEC has developed a seismic hazard software ecosystem that includes traditional probabilistic seismic hazard analysis software (OpenSHA), California velocity models (UCVM), finite difference wave propagation software (AWP–ODC), finite element wave propagation software (Hercules), and physics-based probabilistic seismic hazard software (CyberShake).

Results & Impact

The SCEC team used a physics-based code, RSQSim, to produce long-term (one million+ years) synthetic earthquake catalogs that comprise dates, times, locations, and magnitudes for earthquakes in the California region. After using RSQSim earthquake catalogs to estimate California seismic hazards, SCEC compared them to hazard estimates from traditional, empirically derived California Earthquake Rupture forecasts. These comparisons show that the RSQSim model can replicate seismic hazard estimates with far fewer statistical assumptions.

Why Blue Waters

Blue Waters is a highly productive research environment due to its variety and large number of computing nodes, the large filesystem, the fast I/O system and network, and the science-friendly system administration policies. Blue Waters makes it possible to perform the large-scale physics-based seismic hazard simulations that can reduce the uncertainties in long-term seismic hazard forecasts and improve risk assessments for critical facilities such as large dams, nuclear power plants, lifelines, and energy transportation networks.

Allocation: NSF PRAC/3630 Knh
PI: Leigh Orf
University of Wisconsin-Madison
Geoscience

SIMULATING THE MOST DEVASTATING THUNDERSTORMS: BIG TORNADOES AND BIG DATA
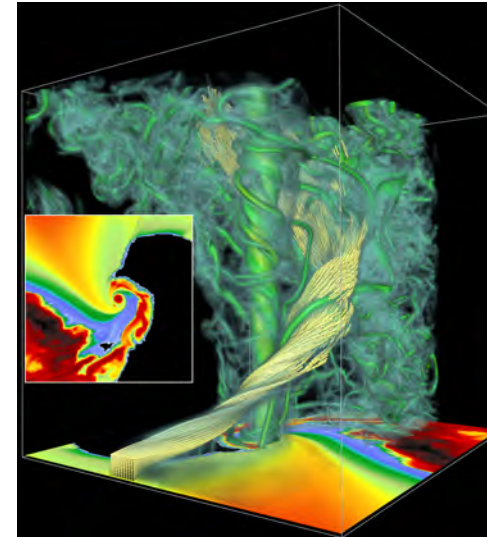
The streamwise vorticity current is comprised of cool air embedded within the storm's forward flank cold pool. It is associated with a drop in pressure1-2 km above ground that results in a very strong low-level updraft that appears to initiate and maintain a long-lived EF5-strength tornado.

Research Challenge

This project's research aims to understand the inner workings of the most powerful tornado-producing thunderstorms in order to better forecast their behavior. The long-term goal is to provide accurate forecasts of these types of storms (as well as less damaging storms) in order to provide ample time for the public to find shelter, and to reduce the high tornado-warning false alarm rate (about 70%) that currently plagues the National Weather Service in the United States.

Methods & Codes

- CM1 cloud model developed at the National Center for Atmospheric Research, a model designed from the ground up to run efficiently on massively parallel distributed memory supercomputers.

- The PI rewrote the I/O driver in order to facilitate frequent saving of large amounts of output, resulting in a file system (LOFS) and associated tools to read back and convert data to widely supported formats such as netCDF.
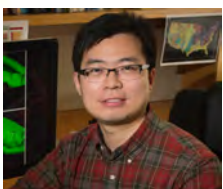
Results & Impact

- These breakthrough supercell thunderstorm simulations conducted on Blue Waters are the first of their kind, containing devastating tornadoes lasting over an hour and a half.

- The identification of the streamwise vorticity current (SVC) in model data has had a big impact in the field of mesoscale meteorology. Efforts are underway to identify the SVC in field studies of supercells.

Why Blue Waters

This project's most modest simulations (30-meter grid spacing) contain over 1.8 billion individual grid elements to resolve the storm, and simulations using 15-meter grid spacing required more than 18 billion elements. The amount of data produced by these simulations is astounding—each simulation producing on the order of 50–100 TB, with the highest-resolution simulations creating closer to half a petabyte of data.

MODELING 4D EARTH EVOLUTION: FROM CONTINENTAL CRATONS TO YELLOWSTONE SUPERVOLCANO

Research Challenge

The challenge geodynamicists face is how to accurately reproduce the various activities within the inaccessible interior of the solid Earth and how to connect them with geological records. To this end, the team's research seeks to answer two questions:

- What causes the most stable portion of continents—cratons—to experience dramatic elevation change and internal deformation?
- What is the ultimate source of heat that fuels the Yellowstone supervolcano?

Schematic view of mantle dynamics below the Western United States. The diverging motion of the hot intruding mantle, corresponding to the Newberry (NB) and Yellowstone (YS) tracks on the surface, is controlled by the sinking Farallon slab (blue) and the retreating Juan de Fuca slab (green).

Methods & Codes

The team uses CitcomS, a community-based finite element code, to simulate a physics-based numerical model of the deep earth. This code, which can be run either forward or backward in time, uses various geodynamic modeling techniques that combine a variety of things such as mantle temperature and viscosity profiles, as well as tectonic plate motion.

Results & Impact

Simulation of the subduction and mantle flow below South America for the past 100 million years revealed that large volumes of the cratonic lithosphere were delaminated into the underlying mantle during the Cretaceous era. This caused the surface to uplift and shed enormous amounts of sediment offshore, leading to thinned crust and deformed lithosphere. This discovery revised the traditional view that the cratonic lithosphere is neutrally buoyant and tectonically stable.

The team also discovered, through modeling of the Yellowstone supervolcano, that most of the heat below the Snake River Plain and Yellowstone caldera originally came from under the Pacific Ocean. This challenges the traditional hypothesis that the Yellowstone supervolcano has been fueled by a deep-mantle plume right below Wyoming.

Why Blue Waters

The enormous amount of data processing and computation required for this work makes Blue Waters the best platform for the team's research.

Allocation: NSF PRAC/9696 Knh
PI: Paul Morin
University of Minnesota
Geoscience

ENHANCED DIGITAL ELEVATION MODEL FOR THE ARCTIC

A shaded relief map of the final release of ArcticDEM.

Research Challenge

Surface topography is among the most fundamental Earth Science data sets, essential to a wide range of research activities, including ice mass-balance, hazard assessment and mitigation, hydrologic modeling, solid earth dynamics, and many others. The National Geospatial-Intelligence Agency, DigitalGlobe, and the Polar Geospatial Center built a near-seamless archive of polar sub-meter stereo imagery that consists of millions of stereo pair images from the Worldview-1, -2, and 3 satellites. Using photogrammetric algorithms, they are able to construct digital elevation models from the stereo pairs, enabling mapping of surface features at the 2-meter scale for the first time.

Methods & Codes

The Surface Extraction from TIN-based Search-space Minimization (SETSM) algorithm, initially designed to extract elevation data over ice sheets, has been refined and optimized to handle stereoscopic imagery over any landcover. After an initial preprocessing step that corrects the source imagery for sensor-specific detector alignment artifacts, SETSM takes the two source images and derives increasingly detailed elevation models using its pyramid-based approach.

Results & Impact

The poles now have better time-dependant topography than almost anywhere else on earth. In addition, The ice on earth has better topography than the land surface on Earth. Our team has produced over 200,000,000 km^2 2m posting topography covering the Arctic over 8 times. The data has been adopted by the Greenland, Iceland and Canada as their national standard elevation data.

Why Blue Waters

No other open research/academic computer has the capacity for this project. Over 17 million node hours (more than ½ billion core hours) were used to process the archive of stereoscopic imagery that was made available. Additionally, the staff at the Blue Waters Project were invaluable in adapting the leadership system to handle the single-node, high-throughput ArcticDEM workflow.

ATMOSPHERIC RIVERS: POTENTIAL DROUGHT BUSTERS
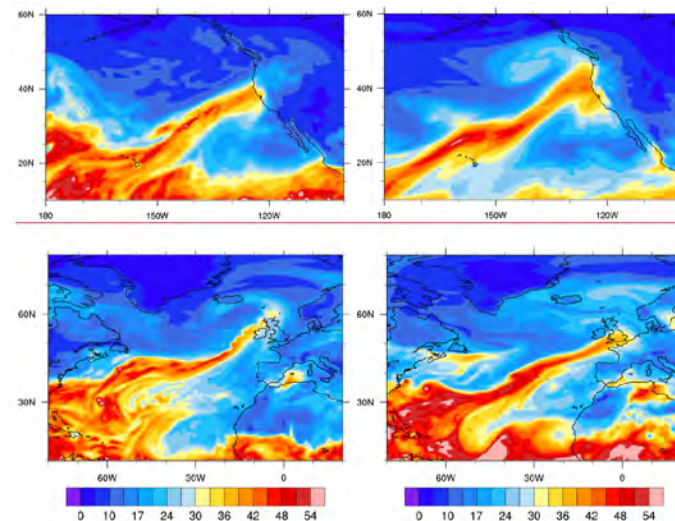
Snapshots of examples of ARs in the North Atlantic (bottom panels) and North Pacific (top panels) when using a standard 1° ocean configuration (right panels) and a high-resolution (0.1°) configuration (left panels). The high-resolution ocean provides significantly more detailed structure to the AR. (Figure courtesy of Christine Shields, NCAR)

Research Challenge

Atmospheric rivers (ARs) are a subset of midlatitude storms that can have a profound regional impact, such as producing heavy rainfall events. This project investigates how ARs on the West Coasts of the United States, United Kingdom, and Iberian Peninsula may change under a future high-emission warming scenario using fully coupled high-resolution atmospheric and oceanic climate simulations.

Methods & Codes

The team used the Community Earth System Model (CESM) for their simulations. A coupled climate model for simulating the earth's climate system, CESM has the ability to simulate Earth's atmosphere, ocean, land surface, sea ice, land ice, and river transport. In this study, the team investigated the representation of ARs and their potential future change using a climate model configuration with a high-resolution atmosphere (0.25°) uncoupled and coupled to both a medium- (1°) and high- (0.1°) resolution ocean. Higher-resolution models match observations better than low-resolution models.

Results & Impact

Rather than focusing on the atmosphere, as most previous AR research has done, the team focused on the ocean's role in AR development, intensity, and structure. They compared simulation results to those of prior, lower-resolution runs, which showed that increasing resolution generally improves the AR count compared to observations. They also investigated changes in ARs that could occur by the mid-21st century under the future high-emission scenario. They found that there would be more or stronger ARs on the California Coast and fewer or weaker ARs in the Pacific Northwest. In the North Atlantic, there would be more or stronger ARs for both the U.K. and the Iberian Peninsula.

Why Blue Waters

Accurately modeling climate processes and projecting future changes requires multiple century-long simulations, including a long, stable preindustrial control, followed by historical and future scenarios. Furthermore, multimember ensembles are needed to quantify and reduce uncertainty. The simulations conducted on Blue Waters this past year consist of a 0.25° atmosphere/land coupled to both a 1° and 0.1° ocean. These high-resolution simulations, necessary for investigation of climate extremes such as ARs, at a minimum require petascale computing resources and cannot be completed without a computational platform like Blue Waters.

DIRECT NUMERICAL SIMULATIONS OF THE RELATIVE MOTION OF HIGH-INERTIA PARTICLES IN ISOTROPIC TURBULENCE

Research Challenge

What are the effects of turbulence on the dispersion, sedimentation, collisional coalescence, and fragmentation of dust grains? The viscous relaxation times, τ_v, of these particles are significantly large, with estimated $St_\eta \sim 10–100$, where $St_\eta = \tau_v/\tau_\eta$ is the Stokes number based on the Kolmogorov time scale τ_η.

Methods & Codes

- Direct Numerical Simulations (DNS) of forced isotropic turbulence through a discrete Fourier-expansion-based pseudospectral method.
- A pseudospectral approach first computes nonlinear terms in physical space and then transforms into spectral space saves computational effort.
- The P3DFFT library transforms between physical and spectral spaces.

Results & Impact

This work can help understand the effects of turbulence on the dispersion, sedimentation, collisional coalescence and fragmentation of dust grains. The "sticking" of dust particles in protoplanetary disks is believed to be the mechanism for planetesimal formation.

Recent progress includes developing a stochastic theory for the relative velocities and positions of high-inertia pairs in forced isotropic turbulence .

Why Blue Waters

This project requires tens of thousands of cores and each DNS run generates several terabytes of data. Due to the CPU time and storage requirements, Blue Waters is the ideal platform. Further, Blue Waters was an invaluable resource in computing key inputs to the stochastic theory, a highly computationally intensive process. This was only possible because of access to Blue Waters, where the code ran on 20,000 cores.

Physics & Engineering

Allocation: BW Professor/245 Knh
PI: Andre Schleife
University of Illinois at Urbana-Champaign
Physics & Engineering

MAGNETO-OPTICAL KERR EFECT OF ANTIFERROMAGNETIC MATERIALS IN EXTERNAL MAGNETIC FIELDS

Research Challenge

Antiferromagnetic materials have interesting properties that can potentially lead to fast switching in future memory devices. Spin compensation makes detecting the spin information of antiferromagnetic materials using optical measurements challenging. Measurements of optical response under external magnetic fields have been developed; however, it is still necessary to understand the fundamental behavior of antiferromagnetic materials better. In order to understand basic principles experiments can benefit from detailed theoretical calculations.

(a) Magnetic unit cell of MnPt; (b) magnetization and energy change of MnPt spin-tilted calculation for corresponding external magnetic fields. Red marker line in (b) presents ground states under corresponding fields and (c) magnon dispersion of MnPt.

Methods & Codes

The Vienna Ab-Initio Simulation Package (VASP) was used for first-principles calculations based on density functional theory to understand the magnetic properties of antiferromagnetic materials. Relaxations of experimental structures were performed in the presence of magnetism and the resulting structures used to compute MOKE signals and Néel temperatures.

Results & Impact

In this project, wavelength-dependent MOKE spectra are computed for MnPt and Fe_2As. The results of magnetic susceptibility and Néel temperature calculations illustrate that this first-principles computational approach can provide reasonable predictions for magnetic properties of antiferromagnetic materials. Thus, our calculations are helpful guidance to understand the behavior of antiferromagnetic materials and to develop advanced magnetic devices.

Why Blue Waters

Computing fully relativistic band structure and its dielectric function with magnetism are demanding processes. Spin-tilted calculations are challenging because noncollinear calculations need more computational resources than collinear calculations. Blue Waters is well suited due to its fast communication and large amount of memory per node and thus provides a unique chance to unveil the unknown properties of antiferromagnetic materials.

Allocation: NSF PRAC/9000 Knh

PI: Sohrab Ismail-Beigi

Yale University

Physics & Engineering
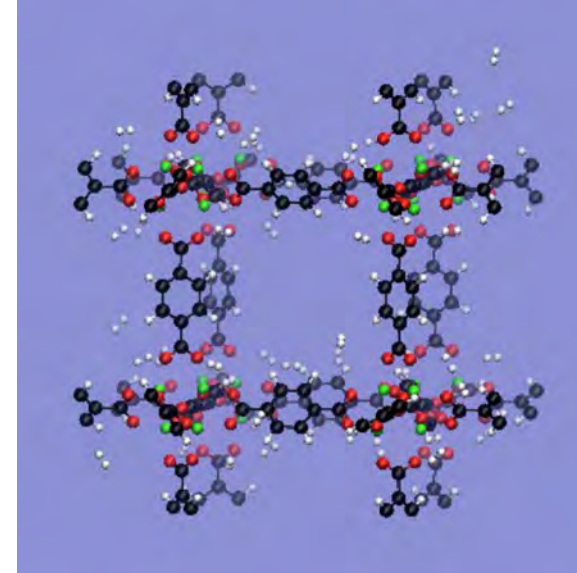
Part of the structure of MOF studied in this project. The black atoms are carbon, the red are oxygen, and the green are zinc. The light grey are the hydrogen that diffuse in the voids.

UNDERSTANDING HYDROGEN STORAGE IN METAL – ORGANIC FRAMEWORKS USING MASSIVELY PARALLEL ELECTRONIC STRUCTURE CALCULATIONS

Research Challenge

Hydrogen, as a fuel, requires efficient storage materials for retaining and releasing hydrogen in large quantities. Metal–organic frameworks (MOFs) are a possible material, and their hydrogen storage potential and microscopic properties need investigation. We performed first-principles quantum-mechanical molecular-dynamics calculations to understand the behavior of hydrogen inside MOFs. The knowledge gained will inform the materials research community on underlying the properties of hydrogen in MOFs, and how we may improve the chemical composition for hydrogen storage.

Methods & Codes

This requires an accurate quantum-mechanical simulation at finite temperature. We employed the Car-Parrinello Molecular Dynamics simulation technique (CPAIMD), which allows the nuclei to move on the Born-Oppenheimer energy surface provided by plane wave-based DFT. By including the electronic degrees of freedom (valence) explicitly, we bypassed force field difficulties.

Results & Impact

The full MOF simulation has a large amount of "vacuum" (voids) in the structure to provide room for hydrogen diffusion, and a large internal surface area for hydrogen adsorption. The smaller size of the mini-MOF allows us to run very long molecular dynamics simulations and benchmark the results carefully. We performed our simulations at 77K and 300K for both systems to understand the effect of temperature on hydrogen diffusion. The simulations show significant diffusion of the H_2 molecules and are statistically meaningful.

Why Blue Waters

Massively parallel electronic structure calculations require tightly coupled computing nodes due to intense communication loads: electron waves are delocalized over the entire system so all parts of the system end up interacting with each other. For the MOF system of interest, the CPAIDM simulations require a massively parallel calculation with many hundreds of nodes.
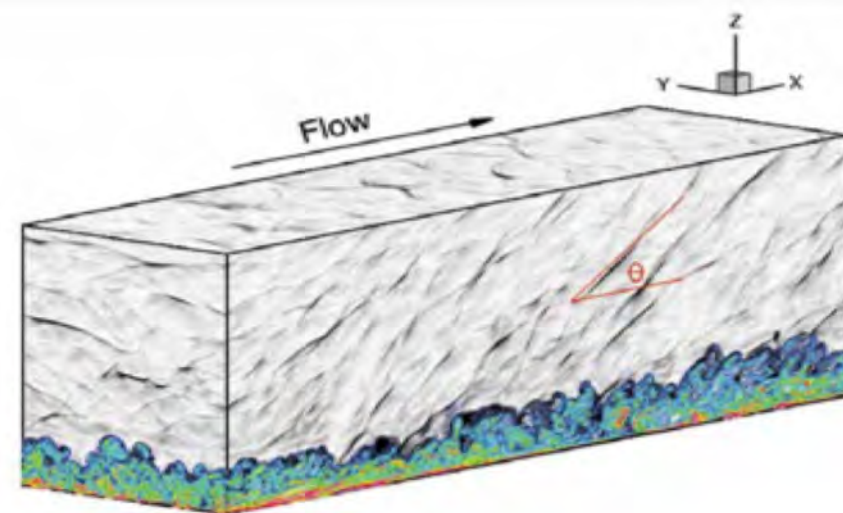
Allocation: NSF PRAC/6900 Knh
PI: Lian Duan
Missouri University of Science and Technology
Physics & Engineering

PRESSURE FLUCTUATIONS INDUCED BY SUPERSONIC TURBULENT BOUNDARY LAYERS

Instantaneous flow visualization for outward-propagating sound waves induced by a Mach 6 turbulent boundary layer. The free-stream acoustic noise is visualized using gray contours of density gradient and the boundary layer is colored by the vorticity magnitude.

Research Challenge

The fluctuating pressure on aerodynamic surfaces induced by near-surface turbulent flow of air plays an important role in vibrational loading during atmospheric reentry. The outward-propagating, free-stream sound waves induced by near-wall air flows are related to the genesis of freestream acoustic noise in high-speed wind tunnels. This research aims to advance fundamental understanding of the surface and free-stream pressure fluctuations induced by supersonic boundary-layer flows for developing physics-based models to predict vibratory loading of reentry vehicles and free-stream acoustic noise in supersonic/hypersonic wind tunnels.

Methods & Codes

Direct Numerical Simulations (HyperWENO) solve compressible Navier–Stokes equations. The convective fluxes are computed using a 7th-order weighted nonoscillatory (WENO) scheme with shock capturing and very low numerical dissipation. A 4th-order central difference scheme is used for the viscous flux. A 3rd-order Runge–Kutta scheme is used for time integration. No-slip conditions are applied for the 3 velocity components; an isothermal condition is used for the temperature. Periodic boundary conditions are used in the spanwise direction.

Why Blue Waters

Direct numerical simulations are used to capture the broadband turbulence field within the boundary layer and the near-field acoustic disturbances radiated by the boundary layer. Extremely fine meshes are required to fully resolve all the turbulence scales in order to obtain the pressure spectra in the high-frequency/large-wave-number range. The simulations need large domain sizes to locate very-large-scale coherent structures. A large number of timesteps are also required for the study of low-frequency behavior of the pressure spectrum. The computational efforts cannot be completed without the world-class capabilities of Blue Waters.

Results & Impact

The work advanced the state-of-the-art of the global pressure field induced by supersonic turbulent boundary layers across a wide range of Mach numbers. This is the first-ever attempt to exploit the advances in HPC to overcome the difficulties in experimental measurements and provide access to both flow and acoustic quantities that are difficult to otherwise obtain. The study led to the 1st successful comparison between numerical predictions and wind tunnel measurements of surface pressure fluctuations underneath hypersonic turbulent boundary layers at Mach 6, 8, and 14.

Allocation: BW Professor/250 Knh

PI: Daniel Bodony

University of Illinois at Urbana-Champaign

Physics & Engineering

RUNNING PLASCOM2 ON BLUE WATERS WITH OPENMP OFFLOADING

Figure 1: Single-node results of PlasCom2 kernels on the Blue Waters system.

Research Challenge

Programming heterogeneous systems is a challenging task, as few programming models support executing code on accelerator devices, leading to the use of specialized solutions such as CUDA, OpenCL, Legion, or Kokkos. This provides the best performance in many cases, as they can often offer support for special device features and good code generation for specific device types. For PlasCom2, the researchers needed an approach which could use the existing code (C++ and FORTRAN) on several accelerator types without having separate implementations for different device types.

Methods & Codes

Based on the offloading support available in recent OpenMP versions (≥4.5), the researchers developed HybridOMP, which is a library to support concurrent execution on host and accelerator devices. For PlasCom2, HybridOMP measures the relative performance of the host and accelerator at startup and determines the best work distribution based on these data.

Results & Impact

OpenMP offloading enables simple and efficient execution of a single code base on different types of devices, with minimal changes to existing code. HybridOMP builds on top of OpenMP offloading to add support for fully heterogeneous execution. Using HybridOMP with PlasCom2 resulted in a speedup of 2.2 times compared to CPU-only execution on a Blue Waters XK7 node. Some of the computationally intensive kernels of PlasCom2 showed a speedup of 5 times as seen in the above figure. These gains remained similar when performing a strong scaling experiment on multiple nodes.

Why Blue Waters

Blue Waters was essential to the research by providing a stable, high-performance platform with easy access to modern accelerators. HybridOMP and its integration into PlasCom2 could be developed directly on Blue Waters to evaluate and compare different implementation possibilities on a real system.
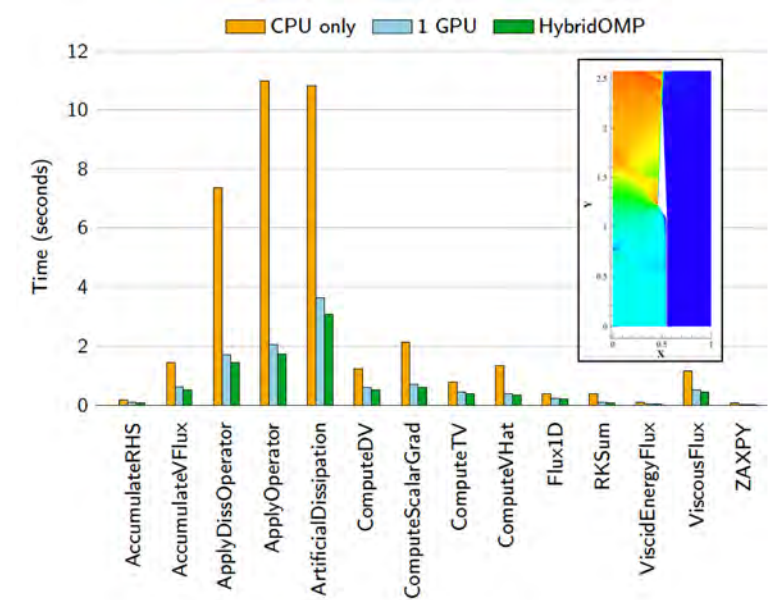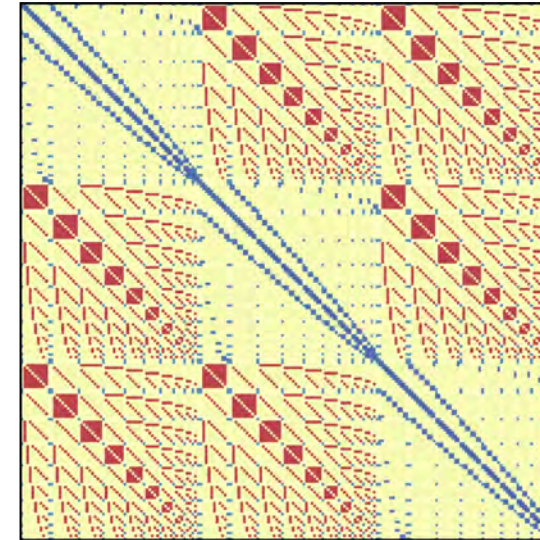
SIMULATING THE EMERGENT PHENOMENA ARISING FROM STRONGLY CORRELATED SYSTEMS

The XXZ0 Hamiltonian has an exponentially degenerate ground-state space. Each of the exponential states in this ground state can be represented as a product state over three spins (here represented as red, blue, and green). Shown is one prototypical three-coloring ground state on the kagome lattice.

Research Challenge

New algorithms are needed in order to understand exotic quantum phenomena. While the rules of quantum mechanics are simple, the resulting phenomena that arise from these rules are difficult to simulate and complicated due to quantum entanglement. Evaluating physically interesting Hamiltonians is exponentially costly due to the needed system size. To avoid this exponential cost we have developed a method that starts with interesting physics encoded in a wave-function and automatically find Hamiltonians that support them. A universal property of complicated quantum systems is the wide array of competing phases that comes from similar Hamiltonians. We seek a unifying explanation for the menagerie of phases by examining a class of materials called frustrated magnets; these are insulating materials whose spin degrees of freedom reside on lattices (grids) such as the triangular or kagome lattice. We numerically probe a transition in a class of physical systems coined the many-body localized phase, a phase where statistical mechanics to break down, between the many-body localized phase and the ergodic phase, where statistical mechanics still operate.

Methods & Codes

- The Eigenstate-to-Hamiltonian construction (EHC), an inverse technique of encoding interesting physics in a wave-function and automatically determining a physically reasonable Hamiltonian.
- A highly parallel exact diagonalization code for XXZ Hamiltonians on a frustrated kagome lattice.
- A novel Shift and Invert MPS algorithm (SIMPS) to compute interior eigenstates of a large, $2^{32} \times 2^{32}$, matrix in the many-body localized phase.

Why Blue Waters

Without computation at the scale of Blue Waters, this project would not have been possible. The examination of many-body localized phase involves studying samples with disorder. To extract any physics requires averaging many thousands of disordered samples that we run in parallel.

Results & Impact

- EHC, opens up an entirely novel approach to solving condensed matter problems and changes the search for interesting physics from a search in the dark to a targeted one. Typical forward approach is exponentially slow: 50 sites is intractable even on Blue Waters. EHC is a quadratic algorithm: **thousands** of sites can be simulated using the inverse approach.
- We have discovered a new Hamiltonian, the XXZ0 point, which has an exponential number of ground states that can be represented as all possible ways of coloring the kagome lattice with three colors.. We showed in a concrete example five explicit phases surrounding the XXZ0 point, including the enigmatic spin-liquid.
- We discovered that many-body localized eigenstates at low temperature can tell that there is an ergodic phase above them at higher temperatures. This allows us to learn about the transition using the many-body localized eigenstates that are easier to probe.

PROPERTIES OF DENSE HYDROGEN

Research Challenge

The properties of hydrogen under extreme conditions of pressure and temperature are of great interest for fundamental research, planetary physics, and energy applications. Hydrogen accounts for much of the visible mass in the universe. The properties of hydrogen and helium are needed for understanding the giant planets, Jupiter and Saturn, but experiments under the relevant conditions are challenging. This research has helped answer many open questions on the properties of hydrogen.

Methods & Codes

The quantum Monte Carlo method used (Coupled Electron–Ion Monte Carlo) works with the full interaction among the electrons and protons and treats both particles fully quantum mechanically.

Results & Impact

The calculations are needed both to validate our computational method and to resolve the different experimental measurements. For progress in the high-pressure community to occur, it is essential to resolve this difference between experiments and computation.

Why Blue Waters

Computational demands from the simultaneous treatment of quantum electrons and protons are very high. Without the size of Blue Waters this would not be possible.

Allocation: BW Professor/60 Knh
PI: Kathryn Huff
University of Illinois at Urbana-Champaign
Physics & Engineering

COUPLED MULTIPHYSICS OF ADVANCED MOLTEN-SALT NUCLEAR REACTORS

The coupled fast neutron flux and temperature in a 3D cuboidal model of a molten-salt reactor. Moltres coupled results have been validated against experimental results from the Molten-Salt Reactor Experiment.
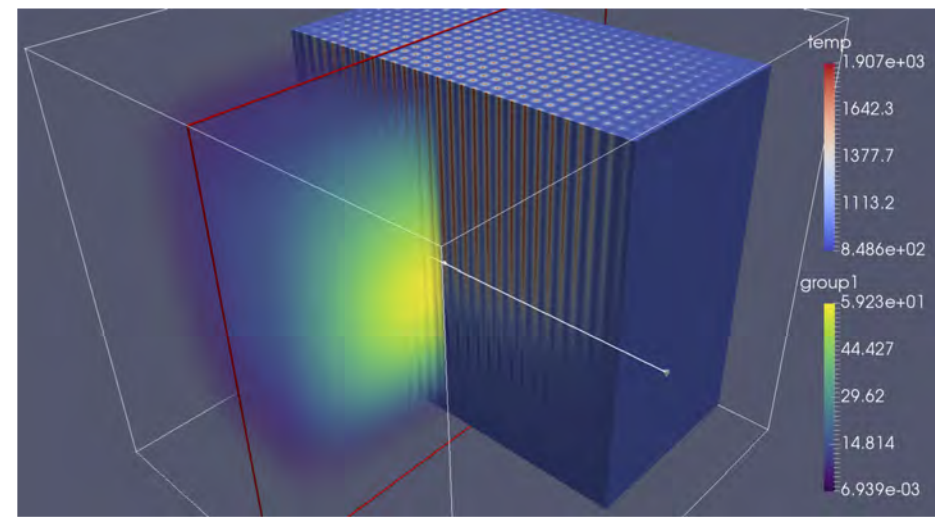
Research Challenge

Nuclear power provides 19% of the total electricity generated in the United States and is our largest source of clean energy. Advanced molten-salt reactor (MSR) technology represents numerous distinct advantages over the light-water reactor (LWR) used today. Current state of the art in advanced nuclear reactor simulation focuses primarily on LWR. This work extends the state of the art by enabling modeling and high-fidelity simulation for MSR designs. This requires development of models and tools for representing unique materials, geometries, and physical phenomena.

Methods & Codes

Two new simulation tools were developed to enable advanced reactor and fuel cycle simulations. First, the finite-element-based physics application, Moltres, couples the thermal–hydraulics and neutronics of molten-salt flow in high-temperature liquid-fueled reactor designs. Second, a Python package for modeling fuel-salt reprocessing, Saltproc. It relies on full-core high-fidelity Monte Carlo simulations to perform depletion computations that require significant computational time and memory.

Why Blue Waters

Accurate simulations at realistic spatial and temporal resolution are only possible with the aid of high-performance computing resources. To assess nuclear reactor performance under a variety of conditions and dynamic transients, myriad 2D and 3D finite element simulations must be conducted. Transient and multiscale simulations, which require greater capability per simulation, are on the horizon for our work. These may occupy up to 100,000 CPU cores at a time.

Results & Impact

Developed in the open, this effort enables both transparency and distributed collaboration on promising nuclear reactor concepts. It can improve designs, help characterize performance, inform reactor safety margins, and enable validation of numerical modeling techniques for unique physics.

MSR technology is exciting and well worth advancing. MSRs are a compelling next-generation nuclear reactor technology as many MSR designs promise improved passive safety, fissile material utilization, recycling flexibility, and power generation responsiveness appropriate for an electric grid involving variable renewable electricity generation.

HIGH-ORDER METHODS FOR TURBULENT TRANSPORT IN ENGINEERING AND GEOSCIENCES

Left: velocity magnitude from a NekNek simulation of a turbulent jet entering a container. Right: two subdomains (orange and black) decomposed from the global domain.

Research Challenge

The goal of this effort is to develop new algorithms for large-scale high-fidelity turbulence simulations for different research problems. The team applies the new apparatus to:
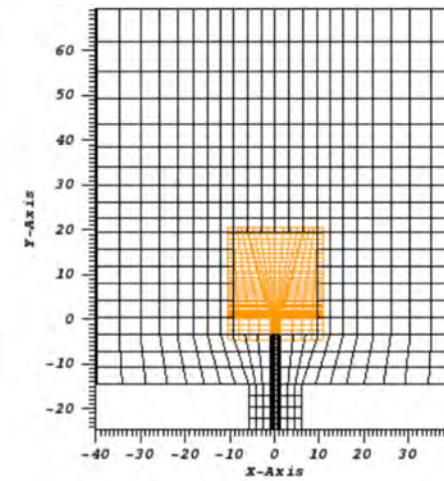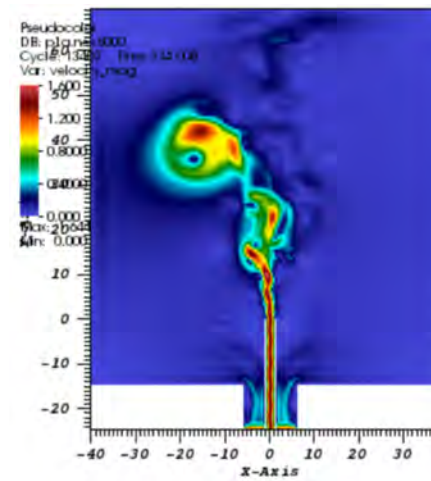
- Better understand how biofilm growth affects the flow hydrodynamics and fine-particle transport at the unprecedent spatial and temporal resolutions.

- Address problems that either have complicated meshes, involve disparate spatial scales, or have moving boundaries that would require the remeshing of standard conforming meshes

Methods & Codes

The turbulence simulations are based on the open-source spectral element code Nek5000. The spectral element method (SEM) is a domain-decomposition approach in which the solution is represented by tensor-product polynomials that cover the entire domain. The team conducted for the first time spectrally accurate direct numerical simulations of a channel with complex natural roughness.

Results & Impact

The team conducted simulations of the flow over biofilm with a bulk Reynolds number of 8,000 with mesh resolutions ranging from 20 million to 200 million grid points. The results show the coherent turbulent structures in the flow caused by the biofilm. The team also used a recently developed method called NekNek to simulate a jet flowing into a tank. The results showed appreciable agreement among simulations with Nek5000 and NekNek, with around 20% reduction in computation cost.

Why Blue Waters

The team conducted simulations using up to 296 million computational points, with the code demonstrating linear speed-up for this problem out to 32,768 MPI ranks. High computation requirements combined with the need for fast turnaround times for the parameter sweep made Blue Water ideal for the task.

Allocation: Director Discretionary/100 knh

PI: Ahmed Taha

University of Illinois at Urbana-Champaign

Physics & Engineering

(a) (b) (c)
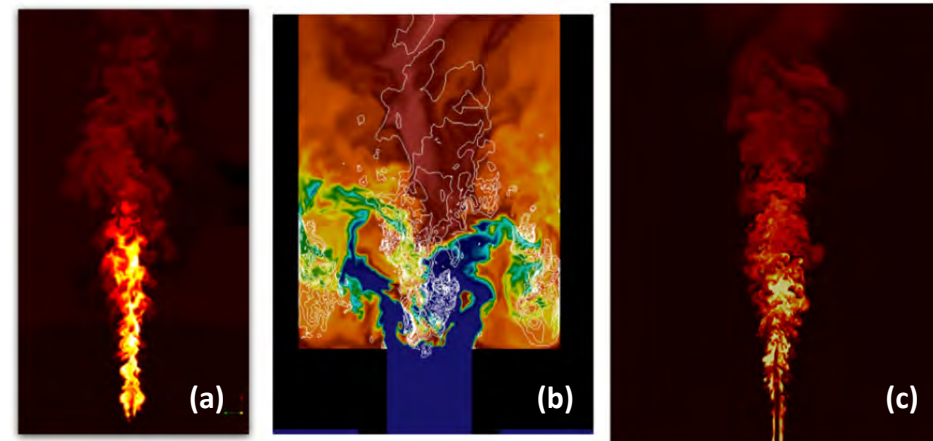
Snapshots of turbulent flames: (a) nonpremixed jet flame, (b) technically premixed jet flame, and (c) nonpremixed piloted jet flame.

HIGH-FIDELITY NUMERICAL SIMULATIONS OF THE REACTING FLOW FIELD OF AN ANNULAR COMBUSTOR MODEL OF AN AIRCRAFT ENGINE

Research Challenge

The team focuses on high-fidelity numerical simulations of aeronautical combustors from turbofan engines using wall-resolved Large-Eddy Simulation. The main challenges associated with modelling of combustors come from the need for full characterization of reacting layers, the near-wall treatment, and the ability to predict pollutant emissions.

Methods & Codes

The project utilizes a multi-physics software package called Alya. It implements finite-element method and has excellent scalability up to 100,000 cores on meshes of up to 4 billion elements. The governing equations describing the reacting flow field are the low Mach number limit of the full Navier-Stokes equations with the energy equation represented by the total enthalpy.

Results & Impact

- Development of a computational framework that can be used to study complex combustion systems based on high-fidelity simulations of turbulent reactive flows.
- Numerical characterization of technically premixed swirling flame.
- Numerical characterization of both piloted and nonpiloted turbulent non-premixed jet flame.

Why Blue Waters

Complete characterization of reacting layers requires an extremely-fine mesh to capture all of the flow details. Besides, achieving convergence at reasonable timescales requires running the code at scale. Using Blue Waters makes conducting the parametric study with such large meshes and complex physical modeling possible in a timely manner.
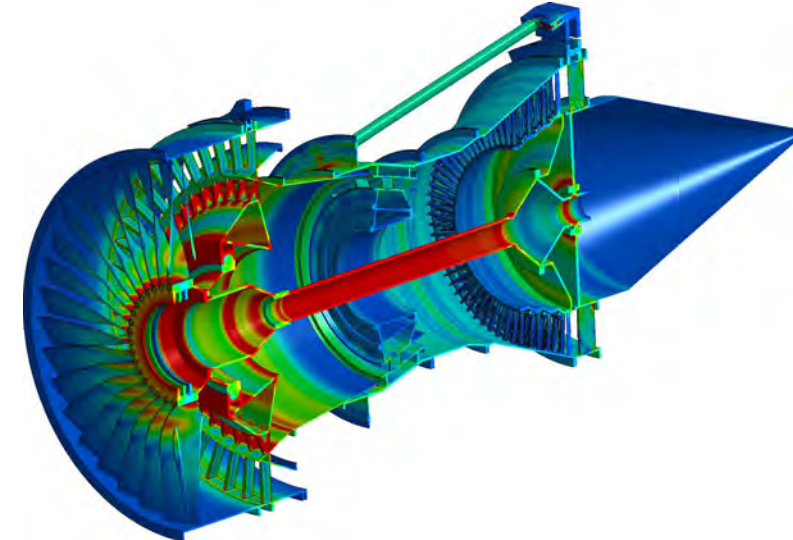
A cross-section of the jet engine model.

IMPROVING VIRTUALLY GUIDED PRODUCT CERTIFICATION WITH IMPLICIT FINITE ELEMENT ANALYSIS AT SCALE

Research Challenge

Modeling and simulation have become increasingly important in both product design and support. Model complexity and size have increased accordingly, leading to longer runtimes for simulation software. The increase in runtime is not linear but, rather, exponential. Research and development of modeling and simulation software are needed to improve its parallel scalability so that larger, more sophisticated models can be run with faster time-to-solution, informing and shortening the design cycle.

Methods & Codes

LS–DYNA is an finite element analysis (FEA) software package used by a wide range of industries. It is used to analyze a diverse set of manufacturing problems such as the simulation of automotive collisions, explosions, and problems with large deformations. These codes provide important features, such as a rich library of elements, contact capabilities, nonlinear material constitutive models, mesh adaptivity, and flexible coupling.

Results & Impact

The research has shown that the implicit finite element method can perform exceedingly well on a petascale high-performance computing platform, thus opening the door for high-fidelity multiphysics modeling in real scale and promoting safer and more energy-efficient cars, airplanes, and other advanced industrial products.

Why Blue Waters

Sparse direct solver algorithms in implicit FEM analysis are both computation- and memory-bound, as well as being communication-bound at large scales. Processing and memory bottlenecks revealed themselves as the number of processors increased by an order of magnitude beyond that familiar to today's developers and users. Many other scientific and engineering codes intended to run on high-performance computing platforms will have similar challenges. Blue Waters is an enabling platform where massively parallel sparse solver technology can be tested and advanced.

Allocation: Director Discretionary/62.5Knh

PI: Sumanta Acharya

Illinois Institute of Technology

Physics & Engineering
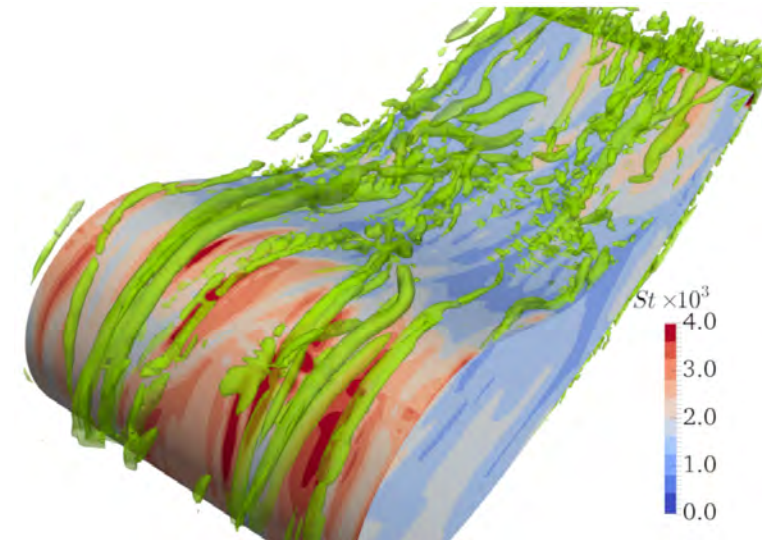
Flow structures and instantaneous Stanton distribution over the turbine vane

HIGH-FIDELITY NUMERICAL SIMULATIONS OF TURBINE VANE HEAT TRANSFER UNDER FREESTREAM TURBULENCE

Research Challenge

Predicting the transition to turbulence typically requires a very fine grid to resolve the smallest scales in turbulence; this, in turn, requires a very large mesh (on the order of 50–150 million nodes for a realistic Reynolds number) and can make the computations time consuming even on large supercomputers. The available transitional models are not universally applicable in different geometries and typically perform poorly. To capture the transition to turbulence, the finest scales available in the fluid flow should be resolved, which requires significant computational power.

Methods & Codes

Multiple Methods and codes were used including: Synthetic Eddy Method, Pressure-Implicit with Splitting of Operators algorithm with the OpenFOAM flow finite volume code, Quadratic Upstream Interpolation for Convective Kinematics method, second-order backward scheme to discretize the temporal derivatives, and WALE subgrid-scale stress model.

Results & Impact

Blue Waters has allowed the Acharya Group to accurately predict the aerodynamics and heat transfer in gas turbines and heat exchange systems, leading to improved fuel efficiency, greater reliability, and lower operating costs in power generation and aircraft transportation.

Why Blue Waters

Blue Waters provides a powerful tool to conduct high-fidelity simulations. Without accessing the computational resources at the scale that Blue Waters provides, performing these simulations is practically impossible. Further, analyzing the temporal behavior of the flow and heat transfer field requires data storage of terabytes of data, which is made possible by both short-and long-term storage solutions provided by Blue Waters.
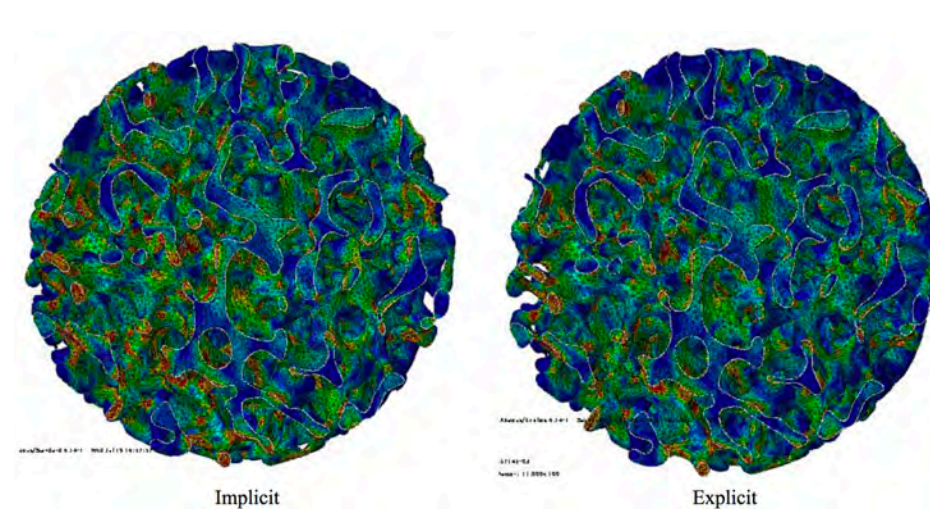
Allocation: Exploratory/50 knh

Ahmed Taha

University of Illinois at Urbana-Champaign

Physics & Engineering

ADVANCED DIGITAL TECHNOLOGY FOR MATERIALS AND MANUFACTURING

Comparison of apparent response resulting from implicit and explicit solvers.

Research Challenge

Bone has a hierarchical architecture spanning from atomistic to macroscopic scales. Osteoporosis is a bone disease characterized by low bone density, which often leads to an increased risk of fractures that mainly occur in trabecular (inner) bone. Trabecular bone is also the primary site for insertion of orthopedic implant systems. Thus, the mechanical properties of trabecular bone are of great clinical and research interest for prediction of age- and disease-related fractures as well as the design of improved implant systems.

Methods & Codes

Modeling of trabecular bone entails a highly nonlinear mechanical behavior along with contacts. As a result, it is of considerable interest to assess the effectiveness and efficiency of an explicit solution method. In this project, researchers used the implicit and explicit solvers of Abaqus to analyze nonlinear micro-Computed Tomography (micro-CT) finite element (FE) models of trabecular bone and compared the performance of the two solvers.

Results & Impact

Results show that by using a similar setup for the model (e.g., element type, loading type, etc.) when using implicit and explicit solvers, there is a perfect match between micro-CT FE model results using implicit and explicit solvers. In addition, the researchers observed that implicit and explicit solvers scale similarly, while the explicit solver performs five times faster.

Why Blue Waters

The team was able to successfully scale their simulations on eight to 12 nodes on Blue Waters with the explicit solver, which significantly reduced computational time. Each of their models have many millions degrees of freedom and nonlinearities, making them impossible to solve without the use of the Blue Waters supercomputer.

Allocation: Exploratory/50 knh

Harley Johnson

University of Illinois at Urbana-Champaign

Physics & Engineering

MODELING MILLIONS OF ATOMS AND THEIR ELECTRONS: TIGHT-BINDING USING A SPECTRAL PROJECTION PURIFICATION METHOD

Research Challenge

To test more efficient methods for computing the electronic structure of systems with millions and tens of millions of atoms. Properties of material domains of this size are important in many different technologies, from microelectronics to medical science to energy technologies. The team focused on properties of dielectric materials of plasma generation, which are important in combustion, materials processing, and catalysis. Traditional electronic structure calculations for these kinds of materials have cubic scaling in N—the number of atoms in the system—making the calculations prohibitively computationally expensive for systems larger than 10^4 atoms.

Methods & Codes

The team developed a new method for accurately and efficiently calculating the electronic properties of materials. Their custom, hybrid code computes a density matrix via combination the of two existing linear scaling methods: the Kernel Polynomial Expansion (KPE) and second-order spectral projection purification (SP2).

Results & Impact

The team has demonstrated their method to be fast, accurate, and efficient in 10-million-atom systems and beyond, which will increase its impact significantly. They plan to apply their method to the benefit of surface and dielectric properties for combustion, catalysis, and materials processing. The team further expects the method could also impact other areas of physics and materials science.

Why Blue Waters

Access to the Blue Waters system made this work possible by permitting studies on a single platform that offered both large parallelism and resources for memory-intensive computations. Depending upon the tuning of the method, it is possible to carry out the entire computation on a single node in a very memory-intensive approach or, by distributing the computation over many processors, it is possible to perform the algorithm in a massively parallel way. The Blue Waters system allowed the team to study the approach and plan for future studies with more optimized tuning.

Allocation: Exploratory/50 knh

Iwona Jasiuk

University of Illinois at Urbana-Champaign

Physics & Engineering
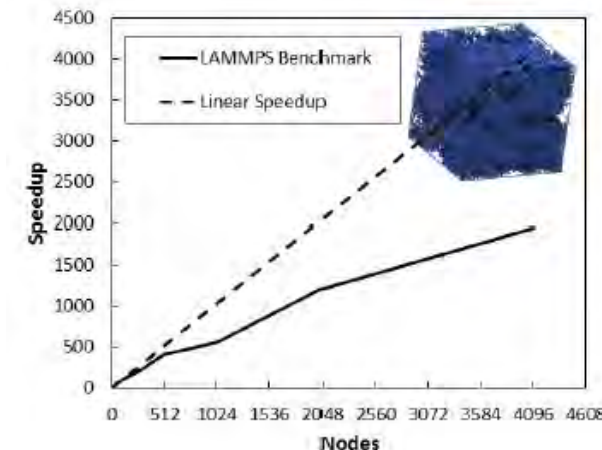
NOVEL METAL-CARBON MATERIALS

Research Challenge

The team used the Blue Waters supercomputer to model novel metal–carbon materials called covetics, which are metals that are combined with carbon in a new way. Experimental results have shown that covetics have superior mechanical, thermal, and electrical properties compared to base metals. However, the role of the added carbon on materials performance is still not well understood. A goal of this project is to improve our understanding of covetics. This project focused on copper covetics.

The Lennard–Jones benchmark runs on Blue Waters XE6 nodes. The parallel speedup curve is plotted for a fixed-sized test case (strong scaling test) with 200 million atoms.

Methods & Codes

The team conducted DFT calculations on covetics by assuming various copper–carbon configurations. They also employed MD modeling, which has not been used previously on covetics. MD allows them to study larger material volumes due to relatively reduced computational cost compared to DFT and is critical in studying experimentally observed structures. They used LAMMPS classical molecular dynamics code, which has a modular design; the community code model allows users to extend the code at the source level.

Results & Impact

Their initial atomistic and continuum models led to the first predictive multiscale models for covetics, to be published here:

Understanding the influence of carbon addition on the corrosion behavior and mechanical properties of Al alloy "covetics". Varnell, J.A.; DiAscro, A.M.; Chen, X.; Gewirth, A.A.; Bakir, M.; Jasiuk, I.; Nilufar, S.; Journal of Materials Science vol: 54, issue 3, 2019, pp. 2668-2679.

Why Blue Waters

The Blue Waters supercomputer and NCSA researchers were essential for this research project. Blue Waters is necessary because such multiscale modeling requires very high computational resources due to the hierarchical structure of covetics.

Allocation: Exploratory/50 knh

Maciej Balajewicz

University of Illinois at Urbana-Champaign

Physics & Engineering

(a) No training   (b) 7k steps   (c) 600k steps   (d) True solution
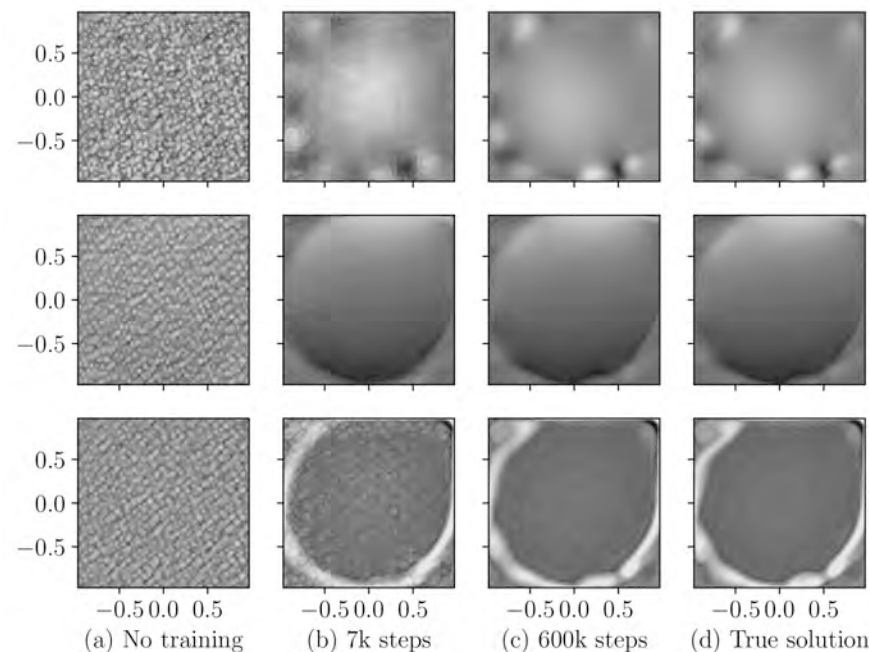
LEARNING LOW-DIMENSIONAL FEATURE DYNAMICS OF TURBULENT FLOWS USING DEEP CONVOLUTIONAL RECURRENT AUTOENCODERS

Research Challenge

For turbulent-flow fluid systems with existing models based on first principles, high-fidelity solutions are possible through direct numerical simulations. However, these generally yield sets of equations with approximately 10^{6-9} degrees of freedom. Even with recent advances in computational power, solving these

Motivated by this problem, this team seeks to develop machine learning approaches to reduce the computational power to model such systems but still predict physical features

Methods & Codes

The code used in this project is written in Python using TensorFlow, Google's open source library for building, training, and serving deep neural network models, which utilizes the CUDA CuDNN deep-learning library for acceleration with NVIDIA GPUs [5]. The model is trained using ADAM, a variant of stochastic gradient descent.

Results & Impact

In this work, we have successfully demonstrated the feasibility of using deep neural network architectures for learning and evolving low-dimensional features of high-dimensional systems through the example of a high-Reynolds-number lid-driven cavity flow.

Why Blue Waters

Training deep neural network models is an inherently data-intensive process. The petascale resources available via Blue Waters, and in particular its large number of GPU-equipped nodes and fast shared parallel storage, have made developing and training deep neural network-based reduced-order models possible.

NUMERICAL SIMULATIONS OF THE INERTIAL COLLAPSE OF INDIVIDUAL GAS BUBBLES NEAR A RIGID SURFACE
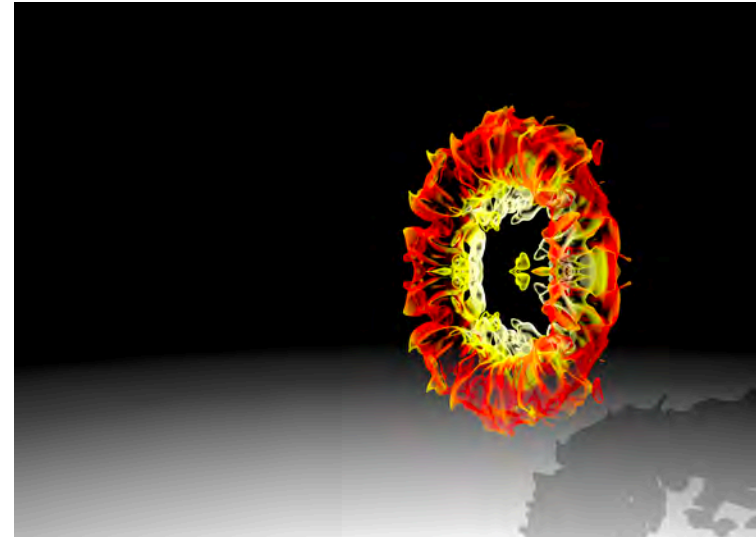
Research Challenge

The research team carries out high-resolution numerical simulations to deepen our knowledge of cavitation-bubble collapse and the damage thereby produced, and help develop strategies to better control cavitation-induced erosion in naval hydrodynamics and for therapeutic ultrasound.

Volume rendering of the bubble shape, colored by its temperature. After the collapse, the bubble takes the form of a hot vortex ring moving in the direction of the jet impact toward the rigid boundary.

Methods & Codes

The research team developed a novel numerical algorithm to solve the three-dimensional compressible Navier–Stokes equations for multiphase flows and designed a spatial scheme to simulate nonspherical dynamics of individual bubbles and the shock waves thereby produced.

Results & Impact

The research team discovered what appears to be a universal scaling of the pressure produced by bubbles collapsing along neighboring solid surfaces, as a function of the pressure driving the collapse and the initial distance of the bubble from the wall. The team has further extended this theory to the temperature produced along a solid surface by a collapsing bubble, and to the collapse of multiple bubbles.

Why Blue Waters

To carry out accurate three-dimensional simulations of bubble collapse in which the small-scale features are adequately resolved, high resolution (up to 2.5 billion grid points) is paramount. Performing such simulations requires a substantial computational power that is difficult to achieve on any other NSF-supported computing machines. A leadership petascale high-performance computing system like Blue Waters is essential for the success of the present study.

Allocation: GLCPC/840 Knh
PI: Venkat Raman
University of Michigan
Physics & Engineering

COMPUTATIONAL MODELING OF ROTATING DETONATION ENGINES

Research Challenge

Study the effect of fuel injection design on the robustness of the detonation process for the next-generation combustor system. This involves full-system simulations of rotating detonation engines (RDEs), which utilize pressure-gain combustion, with an aim of improving the efficiency of power generation and propulsion systems.

Fluid stream trace near the injection system illustrating the complex flow patterns that induce turbulent mixing. Two streams of the oxidizer gases from the left and the fuel gases from the bottom are mixing in the detonation chamber.

Methods & Codes

- Custom compressible flow solver using OpenFOAM and Cantera, UMdetFOAM,
- Monotonic Upwind Scheme for Conservation Laws-based Harten–Lax–van Leer–Contact scheme for discretization of the convective terms
- Second-order Runge–Kutta method for time integration of the Navier–Stokes equations
- Detailed multistep mechanism for hydrogen and air with nine species and 19 steps for chemical source terms

Results & Impact

This research confirms that stable detonations present within the RDE configuration are consistent with experimental observations. This validates the use of meshes created by CAD (Computer-Aided Design) and detailed description of fluid flow. The simulations help the RDE community better understand the physics and optimize this type of combustion system. The team also generated high-fidelity data to assist in the realization of the next-generation combustor.

Why Blue Waters

Blue Waters provided necessary resources to simulate large systems, both several thousand computing nodes to exploit the parallelizability of the solver as well as adequate storage to manage the large amounts of data generated by numerical simulations. Project staff facilitated implementation and multiple job queues were leveraged during deployment.

Allocation: Illinois/550 Knh

PI: Andre Schleife

Co-PI: Alina Kononov

University of Illinois at Urbana-Champaign

Physics & Engineering

ELECTRON DYNAMICS IN ION-IRRADIATED GRAPHENE
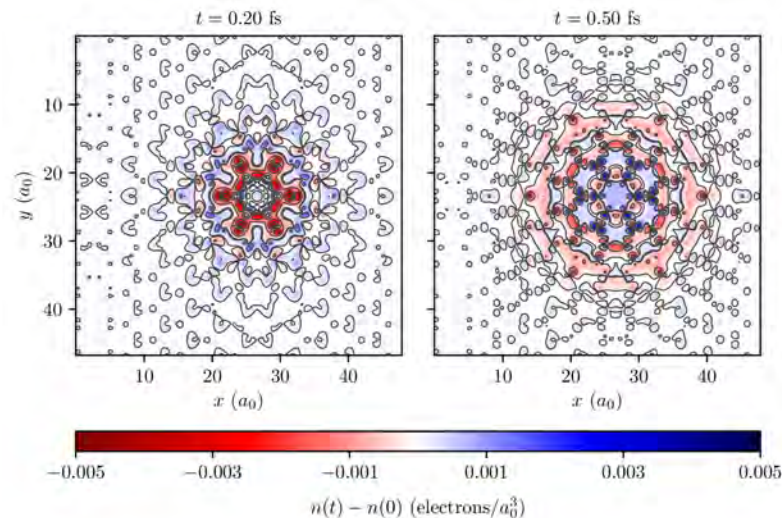
Visualization of local charge in graphene 0.2 and 0.5 femtoseconds after a proton impacts the material. Blue indicates negative charge while red indicates positive charge.

Research Challenge

To understand extremely fast quantum-mechanical processes involving electrons in atomically thin carbon sheets (graphene), which is important for developing radiation-resistant materials and advancing imaging and fabrication techniques for next-generation electronic materials.

Methods & Codes

The team uses Qbox/Qb@ll, their parallel implementation of time-dependent density functional theory. This treats nuclei as classical point charges interacting electrostatically with electrons. Electrons are treated quantum-mechanically; their quantum orbitals are governed by the time-dependent Kohn–Sham equations and are related to the time-dependent electron density. This approach provides accurate information about charge and energy dynamics at time and length scales inaccessible to experiment.

Why Blue Waters

Blue Waters provided the capability to evolve thousands of electrons over thousands of time steps in a large simulation cell. These aspects, which are necessary for accurate predictions, make the simulations computationally expensive; the team was able to compute detailed information about more than 20 setups with different graphene thicknesses and projectile species, charges, velocities, and trajectories.
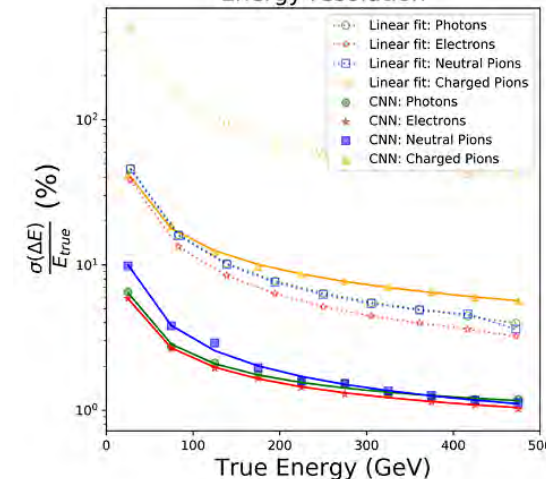
Results & Impact

The team evaluated the suitability of certain parameter combinations for imaging two-dimensional materials and found nearly no change in electron emission for 25–80 keV protons impacting graphene. These findings suggest that proton energies of 25 keV and below are optimal for imaging applications, where high electron emission coupled with low damage to the material are desirable. Such findings inform cutting edge tools for high-resolution microscopy and structure modification of graphene and other two-dimensional materials, whose properties are often sensitive to defects.

Allocation: Illinois/125 Knh

PI: Benjamin Hooberman

University of Illinois at Urbana-Champaign

Physics & Engineering

The relative energy resolution of the four particle types versus the true particle energy.

MACHINE LEARNING FOR HIGH-ENERGY PHYSICS: PARTICLE IDENTIFICATION AND REGRESSION USING DEEP NEURAL NETWORKS

Research Challenge

In 2012, the standard model of particle physics was completed when the Higgs boson particle was discovered using the Large Hadron Collider (LHC). The discovery transformed our understanding of the building blocks of matter and of the fundamental forces. However, the standard model cannot resolve key questions, such as the origin of dark matter, and thus cannot be the final theory of nature. Beyond-the-standard-model scenarios predict the existence of exotic new particles, which may be produced at particle colliders such as LHC. Searching for these particles is the focus of this project.

Methods & Codes

Recent advances in deep learning, along with powerful Graphical Processing Unit (GPU) processors, have made it possible to apply learning networks to large datasets. The researchers simulated signals of electron, photon, charged hadron, and neutral hadron images with the Geant4 simulation toolkit. The images were used to train Deeply-connected Neural Networks (DNN) and Convolutional Neural Networks (CNN). Researchers then used the networks to distinguish signals from electrons and protons from hadron background in the dataset, and to measure particle energies.

Results & Impact

The project team evaluated the performance of Deeply-connected Neural Networks (DNN) and Convolutional Neural Networks (CNN) trained on the particle images, and compared the results to current state-of-the-art algorithms widely used in particle physics. The DNNs and CNNs provided significant improvement compared to conventional methods. These results serve as a first step towards implementing deep learning for particle identification and measurement at the Large Hadron Collider (LHC), and towards potential discovery of exotic new particles predicted by beyond-the-standard-model scenarios.

Why Blue Waters

Optimizing neural network performance requires retraining neural networks hundreds or thousands of times. This is especially challenging for memory-intensive networks. The 4,228 GPU-enabled nodes with 25 TB of GPU accelerator memory available on Blue Waters enabled training of neural networks beyond what had previously been achieved, and allows for detailed investigations of their behavior for both particle physical and general applications.

Allocation: Illinois/75 Knh

PI: Brian G. Thomas

Co-PIs include: SM Cho, H. Yang, SP Vanka

University of Illinois at Urbana-Champaign

Physics & Engineering

MULTIPHASE TURBULENT FLOW MODELING OF GAS INJECTION INTO MOLTEN METAL TO MINIMIZE SURFACE DEFECTS IN CONTINUOUS-CAST STEEL
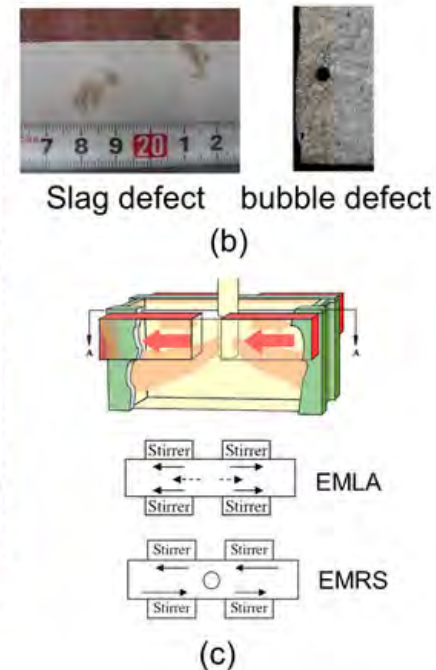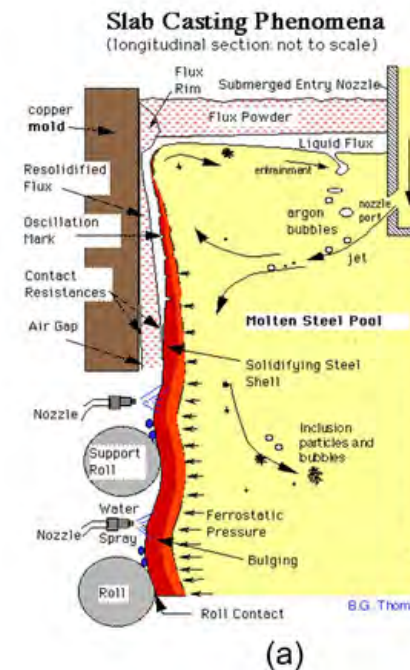
Research Challenge

We develop accurate, comprehensive models of multiphysics phenomena in the continuous casting of steel and apply these models to improve understanding of the mechanisms of the formation of defects in the manufacturing process. This work applies transient multiphase flow simulations to quantify argon gas behavior, bubble size distribution, and bubble transport and capture, which influence bubble defect formation. Further we studied the effect of moving magnetic fields on the flow pattern to investigate ways to reduce defects related to entrapped bubbles and inclusions.

Methods & Codes

We conducted these simulations using the commercial CFD program, ANSYS Fluent High-Performance Computing (HPC) on Blue Waters XE nodes. To calculate bubble transport and capture into solidifying steel shells in the mold, we implemented the LES–DPM–MHD model coupled with the advanced force balance (on each bubble at the solidification front) capture criterion model into Blue Waters XK nodes with the multi-GPU based in-house code CUFLOW.

Results & Impact

Our model was able to simulate realistic argon bubble distributions in the molten steel pool in the continuous caster, which enabled more accurate prediction of bubble defects, including their size and location in the steel product, with and without double-ruler EMBr.

Acknowledgements

Why Blue Waters

Blue Waters enables the simulation of the complex and interrelated transient phenomena on micrometer and millisecond scales in the large volume and long times involved in this large-scale commercial process. We achieved an over three-thousand-times faster calculation with ANSYS FLUENT HPC on Blue Waters XE nodes, compared to an ordinary workstation PC. Furthermore, the multi-GPU in-house CUFLOW codes accomplished a good parallel scalability on Blue Waters XK nodes, showing only 48 hours of wall-clock time for 30 seconds of fully developed LES–MHD flow in a 14-million-cell domain.

Allocation: Illinois/284 Knh

PI: Deborah Levin

University of Illinois at Urbana-Champaign

Physics & Engineering

FULLY THREE-DIMENSIONAL KINETIC SIMULATIONS OF UNSTEADY SHOCK-BOUNDARY LAYER INTERACTIONS

Decomposition of the computational domain held by a single of the 32 processors. Collision and root grids are shown in red and black, respectively. Multiple shocks form as the fluid encounters the wedge and reflected material collides with the incoming fluid forming secondary shocks.
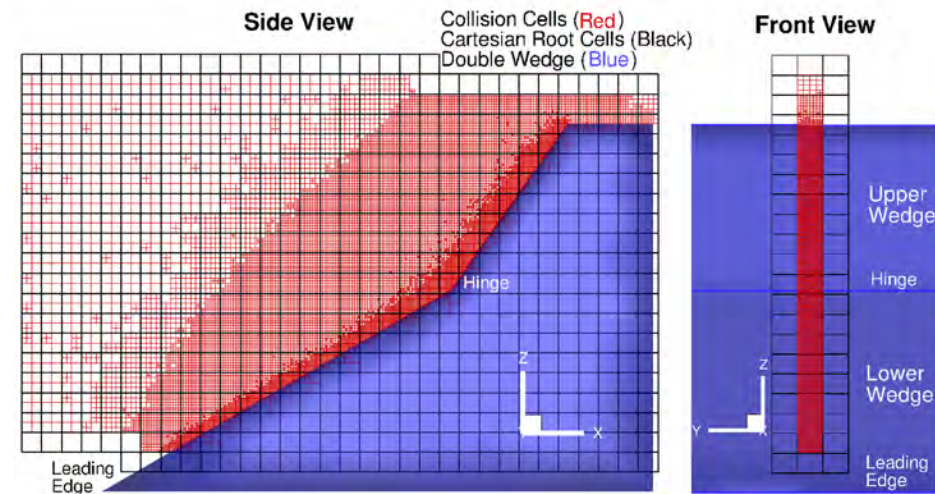
Research Challenge

Air flows encountered during hypersonic flights are characterized by significant gradients in gas density, high temperatures, and a large degree of nonequilibrium. In such flows the shock wave–boundary layer interaction has a significant role in determining aerothermodynamic quantities such as heat transfer, skin friction, and pressure loads. The particle-based Direct Simulation Monte Carlo method provides the highest fidelity in strong shock regions but its high computation requirements require a state-of-the-art resources such as Blue Waters to simulate a large number of particles .

Methods & Codes

The team used the particle-based Direct Simulation Monte Carlo code SUGAR, which is developed by the team. SUGAR uses octree based adaptive mesh refinement for multiscale physics in an unstructured grid. It handles gas-surface interactions efficiently, optimizing memory access for best cache utilization.

Results & Impact

The results for the case presented above were simulated using 20,000 processors, a number that would only be possible on Blue Waters. This simulation provided a detailed understanding of the structure of the shocks formed along the wedge and the exact temperature and density distribution along the wedge.

Why Blue Waters

The time-accurate, large-scale DSMC simulations performed require on the order of 100,000 node-hours, 24 billion computational particles, and 2.2 billion collision cells to reach 1 millisecond of simulated time. Blue Waters is one of the few computer architectures that can host and execute these simulations

INTERFACIAL LOAD TRANSFER MECHANISMS IN CARBON NANOTUBE-POLYMER NANOCOMPOSITES

(a) Close-up views of polymer chains attached to crosslinks at various slip-lengths. (b) Axial force distribution along nanotube with nine crosslinks. (c) Comparison of model predictions for nanotubes versus experimental data.
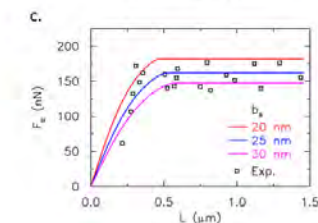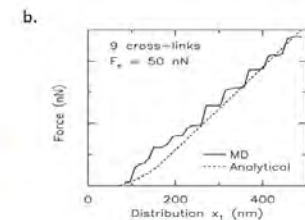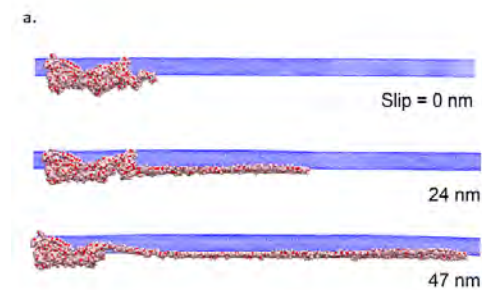
Research Challenge

Carbon nanotubes are highly promising for strength reinforcement in polymer nanocomposites, compared to conventional polymer composites reinforced with micron-scale fillers, which ultimately results in lower component weight. The team used petascale Blue Waters supercomputing resources to provide mechanistic insights into the design of nanocomposites using massively parallel molecular dynamics simulations.

Methods & Codes

The team used molecular dynamics simulations using the open source LAMMPS code in an all-atom simulation of the nanotube-polymer system.

Results & Impact

Using the results of the simulations on Blue Waters the team modified the shear lag model used to describe the behavior of nanotube-polymer composites as a pulling force is a applied. As an outcome, the model can quantify crosslink bond distances for previously performed nanomechanical experiments.

Why Blue Waters

The Blue Waters computational capacities were necessary for several reasons. First, the number of atoms in an all-atom system of ~100 nm- to 500-nm-long nanotubes embedded in polymer matrices ranges from ~0.2 to 10 million. Second, the heating–quenching process adopted to build the polymer configurations required ~3 nanoseconds of simulation time, corresponding to 3 million timesteps for a timestep increment of 1 femtosecond.
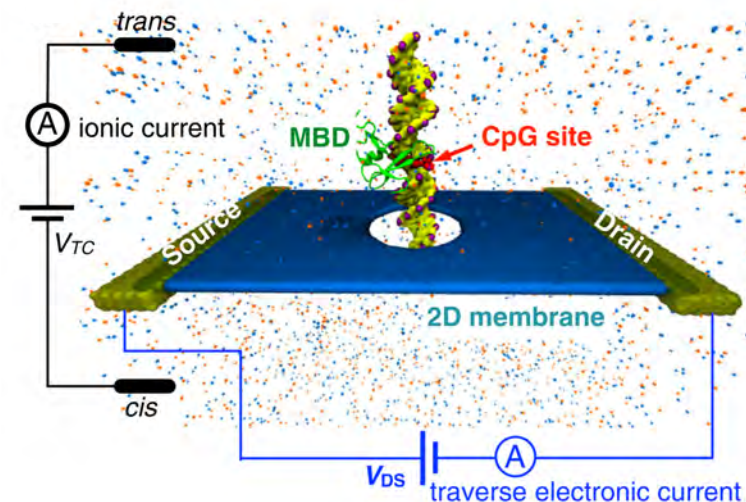
Allocation: Illinois/418 Knh

PI: Jean-Pierre Leburton

University of Illinois at Urbana-Champaign

Physics & Engineering

EPIGENETIC IDENTIFICATION AND MAPPING USING SOLID-STATE 2D NANOPORES

Schematic of the simulated nanopore device. A mDNA-MBD1 complex is being threaded through a nanopore in a 2D material membrane (e.g., graphene or MoS2) embedded in an electrolyte solution. A voltage, VTC, was applied across the membrane to move the mDNA–MBD1 complex, and meanwhile, to induce an ionic current through the pore. Another voltage, VDS, was applied between a source electrode and a drain electrode connecting the 2D material, inducing the flow of a electronic sheet current in the membrane

Research Challenge

Aside from sequencing DNA molecules, the identification of traits of the human genome such as methylation is crucial for diagnosis of epigenetic diseases. Recent experimental evidence of DNA methylation alterations linked to tumorigenesis suggests that DNA methylation plays a major role in causing cancer by silencing key cancer-related genes. Until now, detection and mapping of such DNA methylation patterns using solid-state nanopores have been unsuccessful due to rapid conformational variations generated by thermal fluctuations that result in low signal to noise ratio.

Methods & Codes

Two-step process that first uses molecular dynamics (MD) simulations with the latest NAMD version and then exploits the MD data to calculate the current variations due to DNA translocation through the nanopore via electronic transport modeling. The system is built, visualized, and analyzed using VMD. The protein and DNA are described by the CHARMM22 force field with CMAP corrections and the CHARMM27 force field ,respectively. In the second step, electrostatic and the electronic transport code are written and maintained by the Leburton Group at the University of Illinois at Urbana-Champaign.

Results & Impact

"Noise-free" electronic currents were calculated for all biomarkers. These signatures were compiled into a set of dictionary signals for each of the marker proteins. When a target noisy signal includes the stochastic conformational fluctuations obtained from MD simulations and electronic transport calculations whose marker protein is unknown is fed into the matched filter, it is correlated with the different dictionary signals to identify the marker-protein type. We anticipate that this algorithm can be extended to the detection of multiple markers attached to the same DNA molecule.

Why Blue Waters

Investigation of the interactions of biomolecules with solid-state materials and acquisition of the electronic response using all-atom MD simulations coupled with electronic transport calculations are only possible with petascale computing resources such as Blue Waters. Our systems are about 500K atoms in size, each requiring multiple MD simulation (NAMD) runs. With NAMD code efficiently deployed on XE/XK nodes to run highly parallel simulations of large biomolecular systems, Blue Waters is well suited for our research needs.

Allocation: Illinois/232.8 Knh

PI: Lucas Wagner

University of Illinois at Urbana-Champaign

Physics & Engineering

QUANTUM MAGNETS AND MODELS
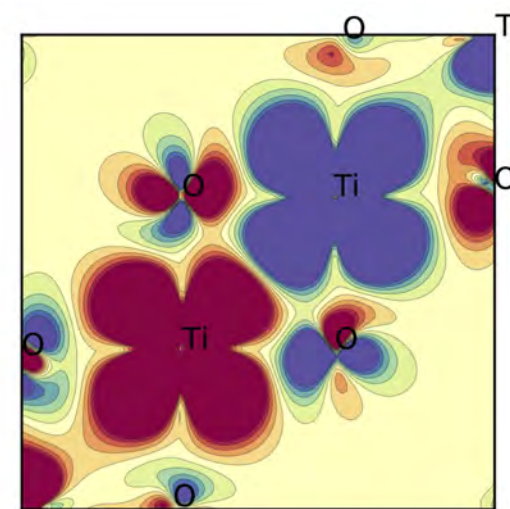
The spin density in $MgTi_2O_4$. Red indicates a net spin-up in a region, while blue indicates a net spin-down. The excitation predicted flips of both spin regions up. The ground state is a quantum superposition of one region up/the other region down with the exchanged configuration.

Research Challenge

Quantum magnetism describes the behavior of single atoms as they act like tiny bar magnets. There is a great deal of interest and science being conducted to look into different materials and how the atoms of specific materials spin hopefully leading to breakthrough discoveries in materials science that would allow for things like room temperature superconductors. Thanks to high-performance computers such as Blue Waters, it is possible to accurately compute the properties of interacting electronic systems using a quantum Monte Carlo method.

Methods & Codes

This project has developed a new technique that maps interacting quantum simulations to coarse-grained models. This requires quantum Monte Carlo simulations that can compute the necessary matrices. To perform the quantum Monte Carlo calculations, the QWalk package was used, a package developed at the University of Illinois at Urbana-Champaign.

Results & Impact

The goal of this work was to produce a prediction of the singlet excitation in the material $MgTi_2O_4$. (A Spinel Oxide). A solid prediction was made with these simulations that will be tested via direct experiment. In the figure, it shows the computed spins, which, when they flip to the same direction, will produce the excitation.

Why Blue Waters

The Blue Waters project staff helped optimize the code used in the calculation. The Blue Waters Symposium (BWS) was an excellent research forum that illuminated modern coding practices such as using the Travis CI based on conversations at the BWS.

HIGH-RESOLUTION MODELING OF TURBULENCE IN WAVE BOUNDARY LAYERS

Right: Turbulent flow structures (colored by velocity magnitude) in an oscillatory flow, Left: flow structures' effect on velocity magnitude (top), vertical component of velocity (middle) and dynamic pressure (bottom).
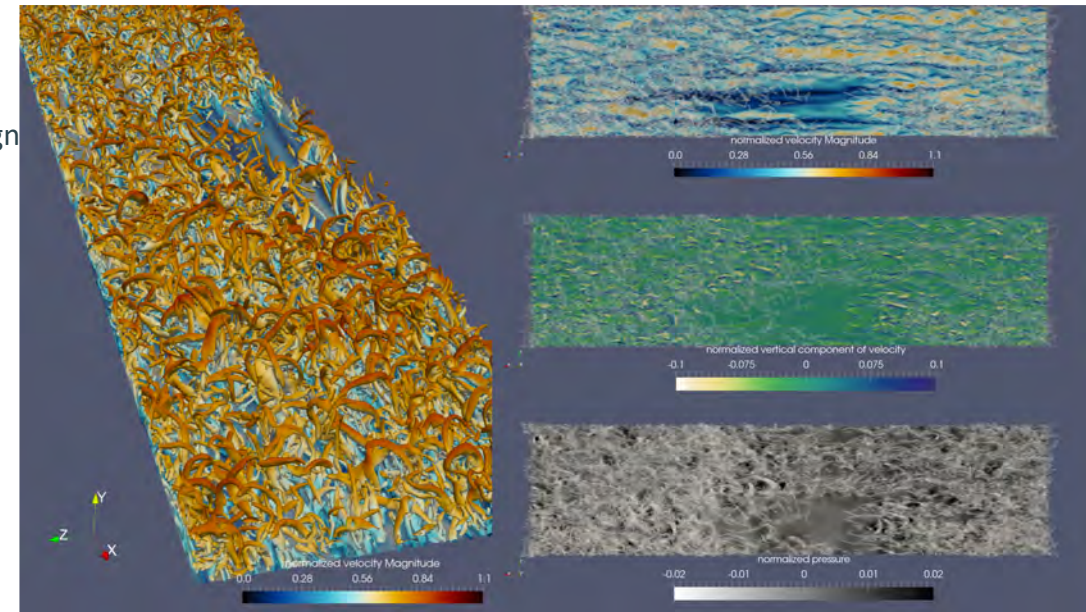
Research Challenge

In the area of physics related to coastal boundary layer flows and sediment transport, this work is the first computational effort to simulate the effect of flow regime and bed characteristics (bed roughness height and bed porosity) on the maximum bed shear stress phase difference compared to the maximum free-stream velocity value.

Methods & Codes

Direct Numerical Simulation model capable of simulating the complex oscillatory boundary layer (OBL) flow and sediment transport using the Spectral Element Method (SEM) framework provided by the highly scalable open-source code Nek5000.

Results & Impact

This work is the first in the literature that explores the effect of turbulence characteristics of the flow and, particularly, the turbulent flow structures, on the phase difference between maximum bed shear stress and free-stream velocity.

The research will lead to a deeper understanding of the interaction of oscillatory turbulent flow, bed shear stress, and sediment mass transport and eventually will lead to the development of new, simplified but accurate models for the analysis and design of engineering systems in coastal and oceanic environments.

Why Blue Waters

Such an analysis pushes the limits of turbulence-resolving numerical studies in terms of the computational resources and the high-performance computing facilities it requires and, thus, it can be materialized only on a petascale supercomputer such as Blue Waters.

Allocation: Illinois/140 Knh

PI: Mark Neubauer

University of Illinois at Urbana-Champaign

Physics & Engineering

(Top) "Images" created by jets of charged particles in Higgs decay signal events and background events that are detected by the ATLAS tracking system. (Bottom) Higgs boson identification accuracy and signal loss as a function of the number of training epochs for a variety of CNN configurations and hyperparameter settings.

ENABLING DISCOVERY AT THE LARGE HADRON COLLIDER THROUGH DATA-INTENSIVE COMPUTATION AND MACHINE LEARNING
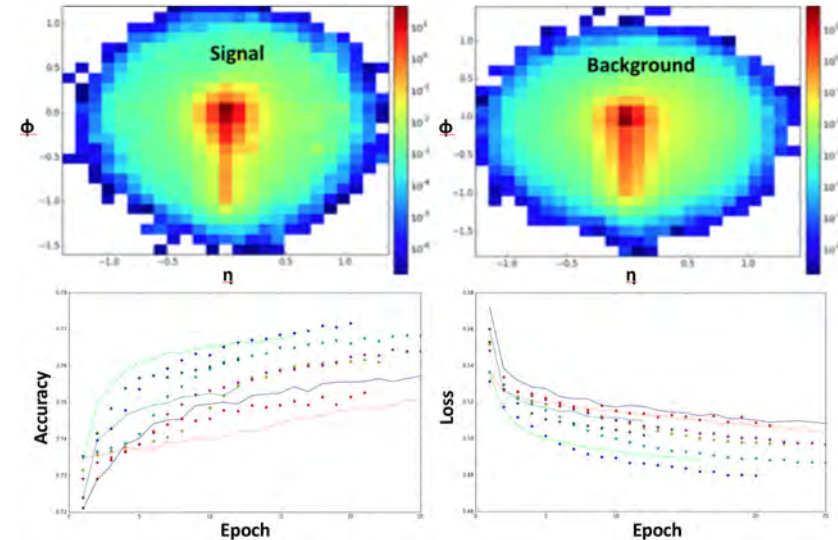
Research Challenge

The goal of particle physics is to understand the universe at its most fundamental level, including the constituents of matter and their interactions. It is an enormous challenge to process, analyze, and share the 50 petabytes of data generated by the Large Hadron Collider (LHC) experiments each year with thousands of physicists around the world. To translate the observed data into insights about fundamental physics, the important quantum mechanical processes and the detector's responses to them need to be simulated to a high level of detail and with a high degree of accuracy.

Methods & Codes

The team has integrated Blue Waters into the production processing environment to simulate and analyze massive amounts of LHC data. The team and collaborators have developed machine learning-based identifiers of boosted Higgs bosons based on (1) a convolutional neural network (CNN) trained using "images" created by jets of charged particles in Higgs decay events, (2) a Recursive NN and (3) a Deep Neural Network (DNN). Researchers have successfully trained these networks using GPUs on Blue Waters via a sequential Keras model with a TensorFlow backend.

Results & Impact

Collision and simulated events processed during the project were made available to the rest of the ATLAS collaboration for use in analysis of the LHC data to improve measurements of the Standard Model and to search for new physics. The techniques show promise in addressing the challenges of boosted Higgs boson identification and improving the sensitivity of new physics searches at the LHC. Of the three techniques, the DNN performed the best and is not being integrated into the ALTLAS code base for broader use.

Why Blue Waters

Blue Waters, as a large CPU and GPU resource with high data-throughput capabilities, greatly facilitated this research. The strong support for containers allowed the team to deploy their science application on Blue Waters nodes. Also, Blue Waters provided a means for a highly parallelized and automated scanning of free parameters in the machine learning configurations and, therefore, rapid optimization of the teams boosted Higgs boson identifier.
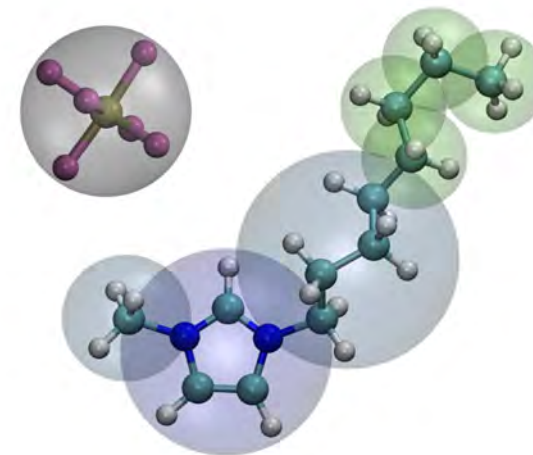
COURSE-GRAINED FORCE FIELD FOR IONIC LIQUIDS

Schematic representation of all-atom and coarse-grained model for $[C_8MIM]^+[PF_6]^-$ with reduced degrees of freedom and similar statistical behavior.

Research Challenge

Ionic liquids have potential applications in energy, healthcare, and nanotechnology. Computational modeling of their behavior, however, is limited by many factors including their slow dynamics, extended range of interactions, and the high computational cost of simulations that are based on all-atom models. The goal of this project was to develop coarse-grained parameters for significantly "cheaper" coarse-grained simulations that could capture complex structure and thermodynamic properties of ionic liquids.

Methods & Codes

Large-scale molecular dynamics simulations were performed using the GROMACS software package. The all-atom model consisted of approximately 100,000 atoms while the coarse-grained model had only 20,000 beads. Coarse-graining was performed using Versatile Object-oriented Toolkit for Coarse-graining Applications making at least 200 iterations for each model to obtain the force fields.

Why Blue Waters

Blue Waters was instrumental to the success of this effort because its enormous computational resources enabled the team to carry out multiple molecular dynamics simulations quickly and efficiently.

Results & Impact

The team extended the relative entropy coarse-graining method with an additional constraint to the objective function as well as long-range electrostatic interactions considerations. The team applied the improved method to develop coarse-grained parameters that were later used to simulate several imidazolium-based ionic liquids. The team has shown that parameters (force fields) are transferable between different thermodynamic states and alkyl chain lengths. This work can pave the way for more accurate simulation of ionic liquids.

Allocation: Illinois/50 Knh

PI: Prashant Jain

Participants: Daniel Dumett Torres & Sudhakar Pamidighantam

University of Illinois at Urbana-Champaign

Physics & Engineering

TOPOLOGIES OF SURFACE ELECTRONIC STATES OF NANOCRYSTALLINE MATERIALS
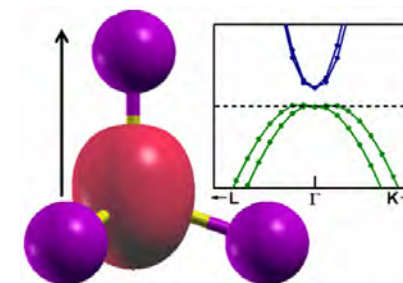
Research Challenge

Engineered nanocrystals have promise in new functional electronic and optical materials. The team uses computational electronic structure investigations to explore the properties of unconventional nanomaterials, such as nanocrystals of $Hg_xCd_{1-x}Se$ alloys. In particular, the team uses Blue Waters to answer the question whether topologically-protected surface states can exist on the surfaces of small nanocrystals and, if not, are there new topologies associated with surface electronic states of nanocrystals.

Calculated band structures show that in the novel wurtzite form of HgSe, symmetry breaking through elongation of Hg-Se bonds results in the opening of a band gap (inset), whereas the conventional zincblende form of HgSe has no band gap. Hg atoms are shown in purple and the electron density of the highest energy valence band is shown in red. Figure published in DOI: 10.1021/acs.chemmater.7b01674 with support of the National Science Foundation (Grant NSF CHE-1455011)

Methods & Codes

For electronic structure calculations, the team uses the open-source Quantum Espresso software suite and the plain-wave density-functional theory method. For studying the effect of nanocrystal size, the team uses projected density of states, charge density maps, surface and bulk band structures obtained from calculations of whole nanocrystals. The computations will reveal whether there is a crystal size regime wherein quantum confinement enhances the band gap while retaining topologically-protected surface states.

Results & Impact

While the team is still a few calculations away from completing this ambitious project, this study has the potential to identify nanostructures with unconventional electronic transport properties of utility in energy storage and information technology. Understanding of the surface states will set the stage for new discoveries about unique electronic behavior on the nanoscale.

Why Blue Waters

The effort required Blue Waters resources due to the scale of bulk slab and full-atom nanocrystal simulations. Numerous calculations, each requiring tens of nodes with a fast interconnect, were required for current investigations, with larger models potentially required in the future. Simulations greatly benefitted from the Blue Waters fast communications hardware.

Allocation: BW Professor/200 Knh
PI: Rafael Tinoco Lopez
University of Illinois at Urbana-Champaign
Physics & Engineering

HIGH-RESOLUTION NUMERICAL SIMULATION OF FLOW AND SEDIMENT TRANSPORT THROUGH AQUATIC VEGETATION
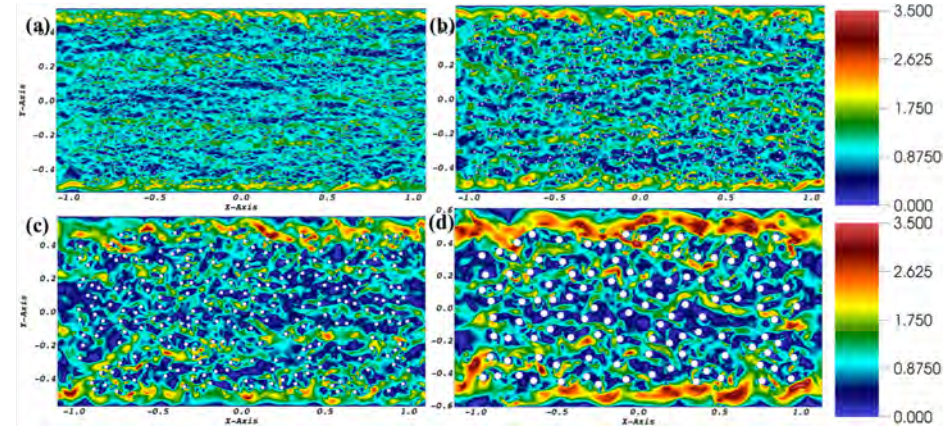
Research Challenge

Aquatic vegetation exerts significant influence in the hydrodynamics of both fluvial and coastal systems. It extracts energy from the flow, which has been found to impede near-bed flow velocities, to modify the structure of the velocity field, and consequently influence sediment erosion and resuspension. The team's study is geared toward increasing understanding of the interactions among vegetation, flow, and sediment.

Instantaneous velocity magnitude for flow through random arrays of idealized aquatic vegetation at Reynolds number ~10,000. The arrays have the same volumetric frontal area but different cylinder diameters: (a) 1/8 inch, (b) 1/4 inch, (c) 1/2 inch, and (d) 1 inch.

Methods & Codes

- Open-source, spectral element-based, high-order incompressible Navier–Stokes solver (Nek5000)
- In the simulations with sediment, sediment transport is modeled under the Eulerian framework using the advection-diffusion equation

Results & Impact

- Vegetation element diameters dictate porosities, with larger diameters increasing both the heterogeneity in the flow and (for the same Reynolds number) the vegetation drag in the transverse direction (2D simulation)
- 3D simulation results show clear evidence of the dynamics of the flow being driven primarily by the vegetation elements, with clear preferential flow-paths through the array of elements

Why Blue Waters

Blue Waters enabled simulations for up to 200 million computational points, with the code scaling strongly up to 16,384 MPI ranks. Without access to petascale high-performance computing, it would be impossible to complete the simulations within a realistic timeframe. In addition, since visualization of a phenomenon is an effective way to understand and explain its mechanics, the team plans to leverage Blue Waters project staff to create animations of the phenomenon using data from the simulations.
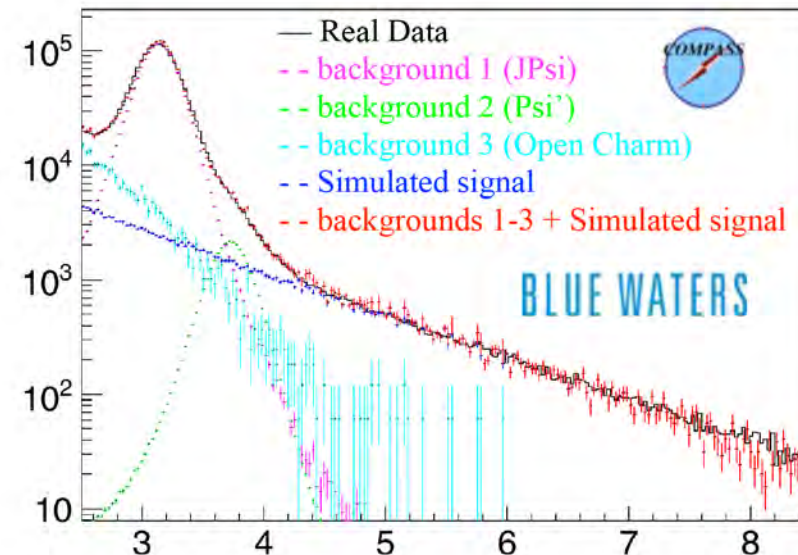
Allocation: NSF PRAC/9440 Knh

PI: Caroline Riedl

University of Illinois at Urbana-Champaign

Physics & Engineering

Simulation of muon pairs reconstructed with the COMPASS spectrometer in comparison with the experimental data.

MAPPING PROTON QUARK STRUCTURE – LOOKING INSIDE THE PROTON: HOW DO QUARKS SPIN?

Research Challenge

- COmmon Muon Proton Apparatus for Structure and Spectroscopy (COMPASS) is a high-energy physics experiment that probes proton substructure by scattering high-energy pion and muon beams off of nuclear targets at CERN in Geneva, Switzerland.

- The experiment explores the momentum and coordinate phase space of quarks inside the proton to shed light on the quark dynamics and provide a critical test of fundamental predictions derived from Quantum Chromo Dynamics (QCD), the quantum field theory describing the nuclear force.

Methods & Codes

- Each triggered event is recorded by the Data AcQuisition system. The COMPASS Reconstruction Analysis Library (CORAL) software performs the transition from raw data information to physical quantities.

- CORAL's function is to reconstruct particle trajectories and momenta as well as the position of vertices.

Results & Impact

- 3 PB of data has been transferred to Blue Waters for data production and analysis

- Blue Waters enables processing an annual COMPASS data set on Blue Waters within five days rather than 50 days on the CERN primary computing clusters

- In addition to raw data processing and physics-level analysis, Blue Waters allows for the detailed simulation of COMPASS detector properties and environmental effects.

Why Blue Waters

- Experimental and Monte Carlo data can be processed 10 to 25 times faster compared to other computing resources available to the COMPASS collaboration.

- Blue Waters enables novel explorations; for example, detector resolutions in kinematic binnings and two-dimensional detector efficiency maps.

- Blue Waters staff provide essential guidance in terms of job flow, load distribution, data transfer, and BW-specific technical features.

Allocation: NSF PRAC/1660 Knh
PI: Gerhard Klimeck
Purdue University
Physics & Engineering
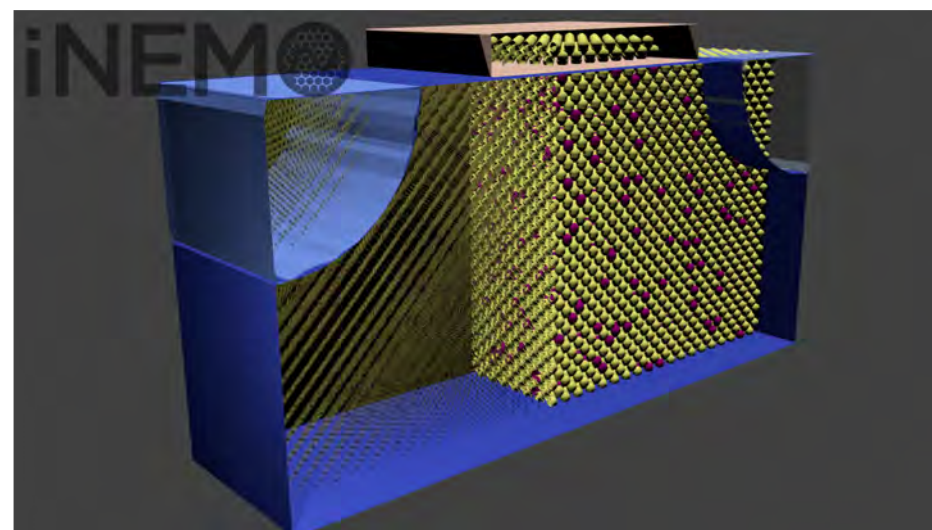
Atomistic representation of a silicon–germanium (SiGe) ultra-thin-body transistor. SiGe is a random alloy with atomistic local disorder. The source and drain are indicated by the rounded shapes. The top gate regulates the electron flow from source to drain.

LEADING FUTURE ELECTRONICS INTO THE NANO REGIME USING QUANTUM ATOMIC SIMULATIONS IN NEMO5

Research Challenge

The United States has always been a world leader in the semiconductor industry with 40% of the worldwide semiconductor device-related patents originating here. The U.S. holds one-third of the global semiconductor device market, which is worth over $300 billion per year. Simultaneously, a relentless reduction in semiconductor size is occurring, with devices expected to be about 5 nanometers long in their critical active region within 10 years. Further improvements in shrinking dimensions can come only through detailed study of device designs, materials, and quantum effects. Complex, fundamental questions remain. NEMO5 was developed by the Institute for NanoElectronic Modeling (iNEMO) at Purdue University to address these fundamental issues. NEMO5 enables basic engineering, physics, and material research, and is used by leading semiconductor firms to design future devices.

Methods & Codes

iNEMO's research on Blue Waters encompasses multiphysics atomistic quantum simulations implemented in NEMO5. The needed physics vary from one device to another, but all can be simulated inside NEMO5. Some examples are:

- Quantum transport simulations (for transistors) using an approach employing semi empirical tight-binding methods.
- Multiscale quantum transport modeling that represent regions of high- and low-electron densities
- Atomic basis sets mapped from established, fundamental methods, and resulting in models that can represent realistically extended devices

Results & Impact

Aggressive semiconductor downscaling into a countable number of atoms in the critical dimensions makes atomistic simulations necessary pathfinders in the quantum regime. NEMO5 is designed to comprehend the critical multiscale, multiphysics phenomena for nanoscale technology through efficient computational approaches and enables quantitative study of new generations of nanoelectronic devices even beyond transistors.
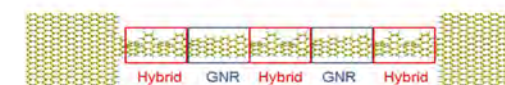
Why Blue Waters

Quantum transport simulations are very computationally expensive and demanding of memory due to the high degree of complexity of the equations used. Blue Waters was used for running such simulations on up to 16,384 cores per simulation. In many cases, the work could not be accomplished in a reasonable amount of time without Blue Waters, and for the larger simulations the work could not be accomplished on other available systems. Blue Waters' staff provide exemplary support and user outreach to guide system usage, help with issues as they arise, and assist with code performance and scaling.

PETASCALE SIMULATIONS OF GRAPHENE NANORIBBON DEVICE STRUCTURES

Calculated I-V curve and structure of a negative differential resistance device based on graphene nanoribbon and graphene nanoribbon precursor

Research Challenge

Creating electrical contact to low-dimensional (low-D) materials is a key to their electronic applications. Low-D metals possessing limited density of states at the Fermi level can enable gate-tunable work functions and contact barriers. Ultra-narrow graphene nanoribbons (GNRs) are example of those. The stability of GNRs against oxidation is critical for their practical applications. The development of atomically precise synthesis of graphitic nanostructures promises a revolution in device design. Controllable fabrication of such devices with atomic precision is the present challenge.

Methods & Codes

The electronic structure calculations use the RMG and PWSCF codes. The quantum transport nonequilibrium Green's function calculations use the localized orbitals and NEGF branches of the RMG code.

Results & Impact

The results indicate that GNRs with variable widths and seamless heterostructures can be realized by using only one molecular precursor. This is different from previous works requiring two different precursors. The study finds that GNR oxidization leads to significantly reduced band gaps of GNRs. This suggests that armchair GNRs can be used as high-temperature oxygen sensors. We designed an experimentally realizable nanoscale device with a practical peak-to-valley current ratio and peak current.

Why Blue Waters

The applications described above require a very large parallel supercomputer with a high-speed interconnect among the nodes (due to frequent exchange of substantial amounts of data among nodes). Each project requires many runs to explore its various scientific issues, with a substantial amount of analysis between the runs. High availability and quick turnaround are thus also very important for timely progress in our research.

COLLABORATIVE RESEARCH: ADVANCING FIRST-PRINCIPLE SYMMETRY-GUIDED NUCLEAR MODELING FOR STUDIES OF NUCLEOSYNTHESIS AND FUNDAMENTAL SYMMETRIES IN NATURE
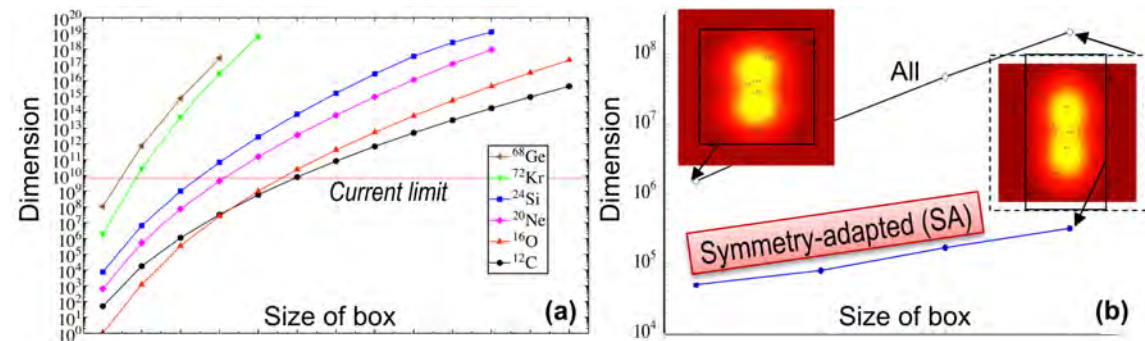

Research Challenge

One of the quintessential open problems in contemporary physics is to design a comprehensive many-body theory for modeling and predicting nuclear structure and reactions. As short-lived nuclei, currently inaccessible to experiment, are often found to be key to understanding processes in extreme environments from stellar explosions to the interior of nuclear reactors, first-principle nuclear models that hold predictive capabilities have had and will have a tremendous impact on advancing our knowledge at the frontiers of multiple branches of physics. This project uses *ab initio* ("from first principles") theory to find a solution to this problem.

Nuclear model space—(a) Explosive growth with increasing particle number and the space (box) in which particles reside (the largest dimension currently attainable is shown by the red horizontal line); and (b) tamed dimensions in the SA framework using symmetries known to dominate the dynamics.

Methods & Codes

The team has developed an innovative *ab initio* nuclear structure approach, dubbed the symmetry-adapted no-core shell model (SA-NCSM), simulated using the computer code "LSU3shell", which embraces the first-principles concept and capitalizes on a new symmetry of the nucleus. The *ab initio* SA-NCSM solves the time-independent Schrödinger equation as a Hamiltonian matrix eigenvalue problem. These theoretical advances, coupled with the computational power of the Blue Waters system, allow them to reach medium-mass nuclei that are inaccessible experimentally and to other *ab initio* methods.

Results & Impact

- Provided the first *ab initio* description of the open-shell ^{19}Ne, ^{20}Ne, ^{24}Ne and ^{24}Si nuclei, with impact on X-ray burst nucleosynthesis and difficult-to-study short-lived nuclei (such as ^{24}Si)
- First principles studies of emergent phenomena in Mg isotopes and their mirror nuclei to provide predictions for deformed and heavy nuclei
- Performed ^{48}Ca and ^{48}Ti studies that impact neutrino experiments which will be conducted at DUNE (Deep Underground Neutrino Experiment)

[Ne – Neon, Si – Silicon, Mg – Magnesium, Ca – Calcium, Ti – Titanium]

Why Blue Waters

Currently, only the Blue Waters system provides the resources required for *ab initio* studies of medium-mass isotopes with high accuracy. To illustrate the level of complexity, applications to medium-mass nuclei require hundreds of exabytes of memory to store the Hamiltonian matrix. The team's current largest production runs utilized efficiently 715,712 concurrent threads running on 22,366 Cray XE6 compute nodes to solve the nuclear eigenvalue problem with a memory footprint of up to 750 TB of data.
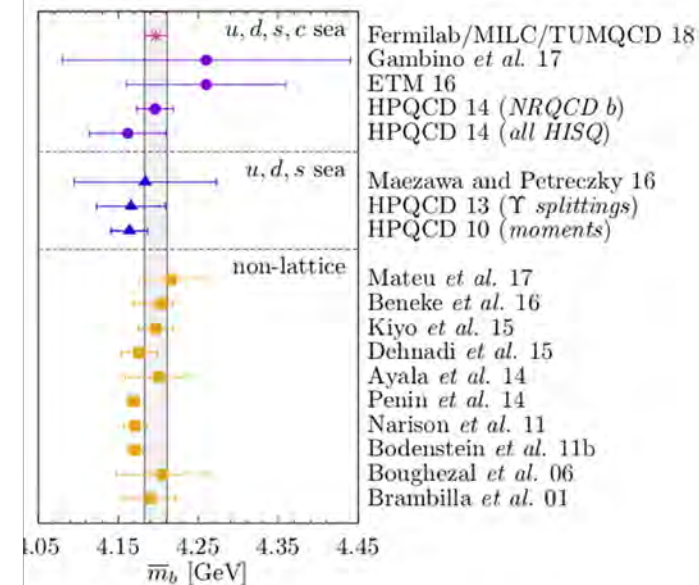
Allocation: NSF PRAC/21780 Knh

PI: Paul Mackenzie

Fermi National Accelerator Laboratory

Physics & Engineering

HIGH ENERGY PHYSICS ON BLUE WATERS

Comparison of our result (magenta burst) for the mass of the bottom quark in the MS-bar renormalization scheme to results from other calculations and from nonlattice methods.

Research Challenge

The standard model of high-energy physics encompasses our current knowledge of the fundamental interactions of nature. The standard model has been enormously successful in explaining a wealth of data over the past 40 years; however, high-energy physicists believe that a more general theory will be needed to explain physics at the shortest distances or highest energies. This project aims at obtaining a deeper understanding of the standard model and at searching for physical phenomena that go beyond it

Methods & Codes

Numerical studies of quantum chromodynamics (QCD) currently use several different formulations of quarks on the lattice, all of which are expected to yield the same results in the continuum limit. The team is using the two formulations most widely employed in the study of high-energy physics: domain wall fermions (DWF) and highly improved staggered quarks (HISQ).

Results & Impact

The HISQ gauge configurations generated under this allocation have been used in the determinations of quark masses and leptonic decay constants of unprecedented precision. In particular, the results for the charm quark mass match the precision of the most accurate previous calculations, and those for the other quark masses and their ratios are the most precise to date. Similarly, the results for the decay constants of D and B mesons are the most precise to date.

Why Blue Waters

Lattice QCD calculations have made major progress in the last few years with a limited number of calculations reaching precision of a fraction of a percent. Such precision is needed to test the standard model and to obtain a quantitative understanding of physical phenomena controlled by the strong interactions. This progress has been enabled by the advent of petascale computers, such as Blue Waters, and could not have been made without them.

Allocation: NSF PRAC/3400 Knh

PI: P.K. Yeung

Georgia Institute of Technology

Physics & Engineering

RESOLUTION EFFECTS AND EXTREME EVENTS IN PETASCALE TURBULENCE SIMULATIONS

Probability density functions of normalized dissipation (red) and enstrophy (blue) in isotropic turbulence, at higher spatial and temporal resolution than in previous work. Dashed lines show very close fits with stretched exponentials. Line in green shows results for twice the dissipation.
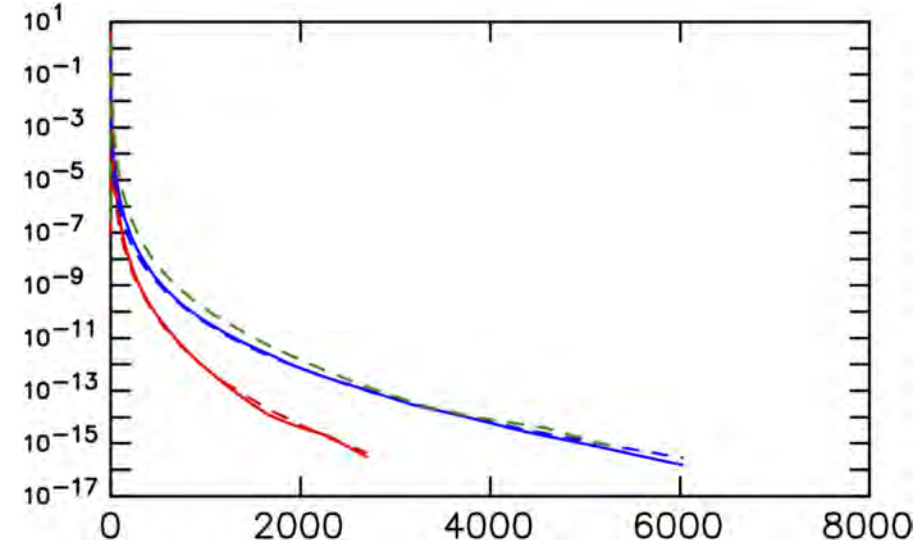
Research Challenge

- Intermittency is a fundamental but still not well-understood property of turbulence at high Reynolds numbers useful for predicting flow patterns in different situations in numerous fields.
- A basic question is how large the fluctuations can be, with what likelihood, and whether the dependence on Reynolds number can be quantified reliably. While attaining a high Reynolds number is always desirable, theoretical models have also suggested a need to resolve the small scales better than is often practiced

Methods & Codes

- Solving the Navier–Stokes equations with the Fourier pseudospectral methods in space and finite differencing in time for, while calculating the probability distribution of fluctuations of energy dissipation rate and enstrophy on the fly, is found to be beneficial for studying high-intensity, small-scale fluctuations that evolve on very short time scales.
- An approach progressively filtering out spectral content at high wavenumbers is developed to identify adverse effects of numerical errors.

Results & Impact

- The observation that dissipation is affected by resolution effects more than enstrophy suggests past statements concerning comparative behaviors of extreme events in these quantities may require revision.
- The conclusion is that the probability density functions (PDFs) of these two variables do not coincide in the range where extreme events occur. However, they do possess a remarkable degree of similarity, with both being well described by stretched-exponential fitting functions differing in only one parameter.

Why Blue Waters

The 8,192³ grid resolution requires access to a world-class machine such as Blue Waters, and the memory capacity of Blue Waters has made the 12,288³ simulation with good time resolution possible.

Allocation: NSF PRAC/4050 Knh

PI: Warren Mori

University of California, Los Angeles

Physics & Engineering

PETASCALE SIMULATION OF HIGH-ENERGY-DENSITY (HED) PLASMAS

Research Challenge

The key research questions are: Can plasma-based acceleration be the basis of new compact accelerators for use at the energy frontier, in medicine, in probing materials, and in novel light sources? Can laser–plasma interactions be controlled or even harnessed in inertial fusion energy-relevant plasmas? What are the collective processes responsible for the formation of shocks in collisionless plasmas? Are collisionless shocks in plasmas responsible for the most energetic particles in the universe?

Spatial distribution of beam intensity from 3D OSIRIS simulation of laser–plasma interactions relevant to inertial confinement fusion. The pattern mimics the effects of optical elements that are present in current fusion experiments and is important for achieving quantitative agreement among simulations and experiments.

Methods & Codes

The UCLA Plasma Simulation Group and its collaborators at IST in Portugal maintain a large number of PIC codes, including OSIRIS, QuickPIC, and UPIC.These codes are all developed locally, share many of the same algorithms and data structures, and have been optimized for heterogeneous leadership-class supercomputers such as Blue Waters.

Results & Impact

In the past 12 months, the results of these petascale simulations have appeared in high-impact journals such as Physics of Plasmas and Physical Review Accelerators and Beams.

Why Blue Waters

The system has continued to provide a very stable high-performance platform for the study of kinetic effects in high-energy-density plasmas.This stability has allowed the team to perform a large number of petascale simulations that will help experimentalists produce brighter X-ray sources using X-FEL (using LWFAs) and produce higher-yield targets in inertial fusion experiments.

Proton migration from the surface to bulk of perovskite structure.

MECHANISTIC INSIGHTS INTO HYDRATION OF SOLID OXIDES

Research Challenge

- Solid oxide electrolysis cells (SOEC) operating at a high temperature will help reduce cost and enhance efficiency for hydrogen production as an alternative energy source.
- SOECs with proton-conducting electrolytes such as Y-doped $SrZrO_3$ (SZY) can produce pure and dry hydrogen.
- The design and development of solid oxide electrolytes with sufficient stability and enhanced conductivity for SOEC are still a challenge.

Methods & Codes

- Density functional theory (DFT) calculations are used to investigate the hydration behavior of solid oxides and performed using the Vienna Ab initio Simulation Package (VASP)
- The cutoff energy for the plane-wave basis set was 500 eV for all calculations. All the calculations are spin polarized. Bader population analysis is performed to calculate the atomic charges.

Results & Impact

- The project investigated the hydration processes in two solid oxides with a single-crystal structure—a proton-conducting SZY and an oxide ion-conducting YSZ.
- Results - both the SrO (001) surface of SZY and (111) surface of YSZ are favorable for the adsorption of water. The structural minimization and charge analyses indicate that the water molecule is adsorbed as two hydroxide ions on the surface improving the efficiency of the process

Why Blue Waters

- Large-scale ab initio simulations are needed to obtain the dynamics of hydration in solid oxides.
- For these calculations, using multiple nodes for each job achieves the best performance and the highest productivity of the science team.
- The computational productivity improvement is attributed to the power of Blue Waters and the support of project staff.

Computer Science & Engineering

ALGORITHMS FOR EXTREME SCALE SYSTEMS
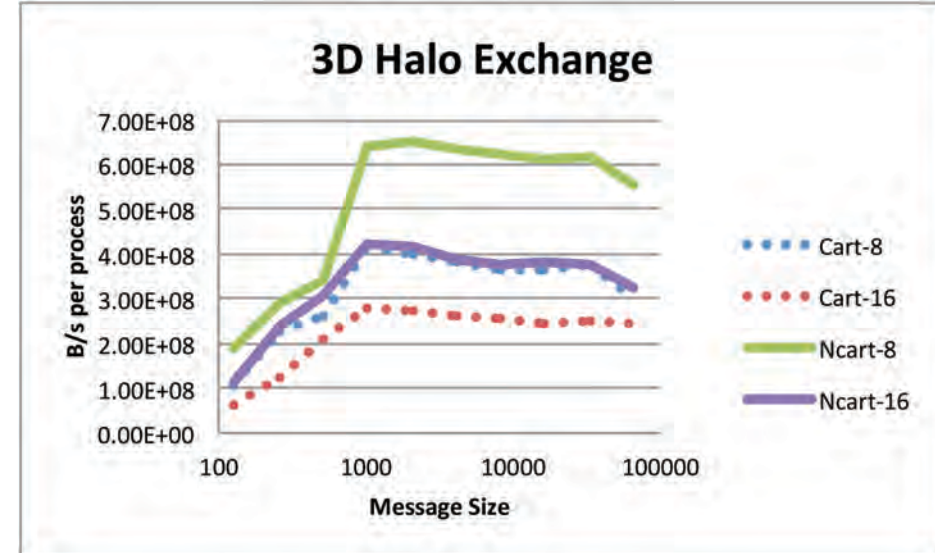

Research Challenge

This work directly benefits applications on large scale multicore systems that are communication intensive, and include collective all-reduce operations that limit scalability. Our approach begins with developing a performance model that captures the key aspects of the intra- and internode communication costs, and uses that model to inform the development of new algorithms. This approach has also yielded improved parallel I/O routines and a better implementation of process placement that improves the performance of applications.

Significant performance benefit achieved on large scale multicore systems from an alternative implementation of MPI_Cart_create that was developed based on a new performance model.

Methods & Codes

- Developed a new communication model for multicore nodes

Results & Impact

- Better understanding of Krylov solver performance at scale, more accurate performance models, and optimizations of the solvers

Why Blue Waters

Blue Waters provides one of the few available environments where such large-scale experiments can be run. In addition, Blue Waters provides a highly capable I/O system, which we used in developing improved approaches to extreme-scale I/O.

Allocation: Exploratory/60 knh

Luke Olson

University of Illinois at Urbana-Champaign

Computer Science & Engineering

SCALING ELLIPTIC SOLVERS VIA DATA DISTRIBUTION

Weak scaling of a 3D elliptic problem on Blue Waters with a per-core size of 5M+ and a total problem size of over 660 billion points at the largest scale.

Research Challenge

Structured multilevel methods are a highly effective and common tool in solving elliptic partial differential equations across a range of applications areas. Scalability of these methods is limited, particularly at high core counts. The focus of this work is on the development of a multilevel method that redistributes data on coarse levels in the solver in order to optimally extend the parallel scalability of the solver.

Methods & Codes

To test a predictive performance model for scaling the solver to 100,000 cores, while the baseline solver is limited to 4,000 cores a method is used to:

- Agglomerate processors into groups of coarse tasks;
- Gather or collect data for processors in each block; and
- Continue cycling redundantly with redistributed data.

The code used and tested for this work is the Cedar Framework (https://github.com/cedar-framework/cedar).

Why Blue Waters

Blue Waters was instrumental in developing accurate performance models and in testing the scalability of the methods. It helped to extend the method and the code to new scales.

Results & Impact

This work contributed to fundamental advances in structured solvers.

- Testing of a new scalable solver - leading to a robust and efficient method for a range of structured elliptic problems.
- Development of an efficient halo exchange suitable for heterogeneous systems.
- Modeling led to new insights on parallel performance.

Allocation: Exploratory/48 knh
Matthew West
University of Illinois at Urbana-Champaign
Computer Science & Engineering

Optimization efficiency of different algorithms for hyperparameter-tuning problems. The total time on the horizontal axis is the actual elapsed time. Our ProSRS algorithm is significantly faster (more efficient) than the other algorithms.

AN EFFICIENT OPTIMIZATION ALGORITHM FOR AUTOMATIC TUNING OF MACHINE-LEARNING MODELS
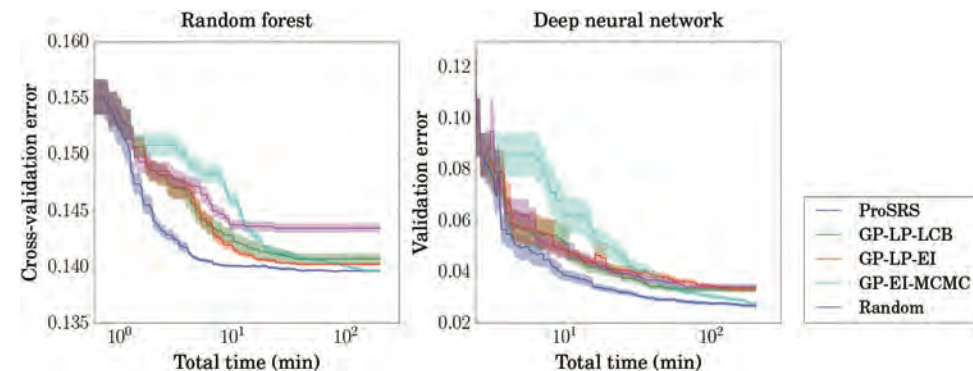
Research Challenge

Machine-learning algorithms have a set of tunable configuration parameters, known as hyperparameters which generally have a huge impact on the performance of a machine-learning algorithm. The difference between poor and good hyperparameter settings can mean the difference between a useless model and state-of-the-art performance. Bayesian optimization is a popular method for tuning hyperparameters, but it has a high computational cost, limiting the range of tuning. This project is focused on developing a more efficient optimization algorithm for hyperparameter tuning.

Methods & Codes

Developed an efficient optimization algorithm (ProSRS) for hyperparameter tuning that uses radial basis functions as well as a novel "zoom strategy" to improve efficiency of the algorithm. Codes are structured in a master-worker configuration – the optimization algorithm (master) proposes multiple sets of hyperparameters to a batch of workers – each works in parallel to train a machine-learning model. The algorithm is implemented in Python with common Python libraries (Numpy, Scipy, Scikit-learn and Mpi4py).

Results & Impact

Compared their algorithm, ProSRS, to three state-of-the-art parallel Bayesian optimization algorithms: GP-EI-MCMC , GP-LP with LCB, and EI acquisition functions and found that the ProSRS algorithm performed the best. The ProSRS algorithm shows superior optimization performance with significantly lower cost meaning it is suitable for a wider range of hyperparameter-tuning problems, not just very expensive tuning problems.

Why Blue Waters

To perform a full numerical experiment, the team needs to compound the base machine-learning computation with the number of tested algorithms, the number of hyperparameter-tuning problems, and the number of repeated runs for each algorithm and each problem. Therefore, this project demands large-scale computation that only Blue Waters can provide.

Allocation: Exploratory/50 knh
Ravishankar Iyer
University of Illinois at Urbana-Champaign
Computer Science & Engineering

Characterization and Diagnosis workflow for interconnection-networks. The data set size highlighted in red above each block represents the size of the data set collected from Blue Waters.

HOLLISTIC MONITORING AND DIAGNOSIS OF LARGE-SCALE APPLICATION PERFORMANCE DEGRADATION
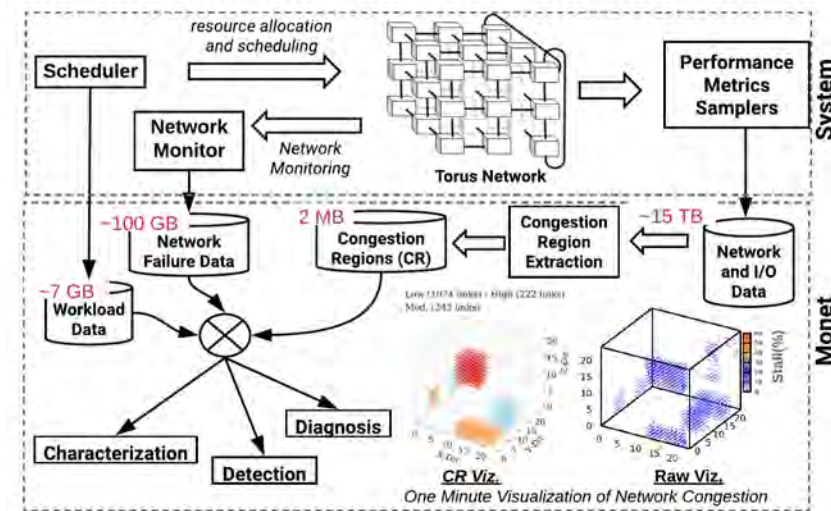
Research Challenge

This project presents a methodology for characterizing congestion in large-scale high-speed 3D Torus networks. Production characterizations of congestion manifesting as hot spots can be difficult, since they require systemwide, coherent data on the state of the network. In addition, data must be collected at fidelities necessary to capture the relevant phenomena. Runtime evaluations that identify localized congestion and assess congestion duration can be used to trigger mitigation responses, such as resource scheduling, placement decisions, and dynamic application reconfiguration.

Methods & Codes

This project developed *Monet*, a generic framework for supporting congestion characterization and diagnosis in HPC systems. *Monet* aggregates and analyzes the network and I/O, resilience, and workload data. The network stall counters are used in the extraction of congestion regions. The identified congestion regions are then combined with other data sets (workload data, network failure data, and network performance data) to enable detection, diagnosis, and characterization of network congestion.

Results & Impact

The project team has shown an example use case in which their analysis methodologies and framework, *Monet*, were used to detect and correctly diagnose the congestion for an execution of the Enzo application, which was reading from the file system at the time of detection. The project-developed tool generates evidence for system managers and users by producing timeseries data and statistical distributions of stall and traffic characteristics for the implicated application.

Why Blue Waters

Blue Waters is one of the few open-science capacity systems that provides a test bed for scaling computations to tens or hundreds of thousands of cores on CPUs and GPUs. It also enables the study of failures and degradations of applications in production petascale systems with its unique mix of XE6 and XK7 nodes. This allows understanding the performance–fault-tolerance continuum in HPC systems by enabling the investigation of application-level designs for mixed CPU and GPU node systems, and fault isolation in system components to mitigate failures at the application level.

Allocation: Exploratory/40 knh

PI: Roy H. Campbell

University of Illinois at Urbana-Champaign

Computer Science & Engineering

ACCELERATING DEEP NEURAL NETWORK TRAINING WITH MODEL-AWARE NETWORK SCHEDULING

Communication to computation vs. communication overlap ratio achieved in TensorFlow showing that the Caramel implemtation outperforms commonly used Horovod and Parameter Server(PS) implementations.
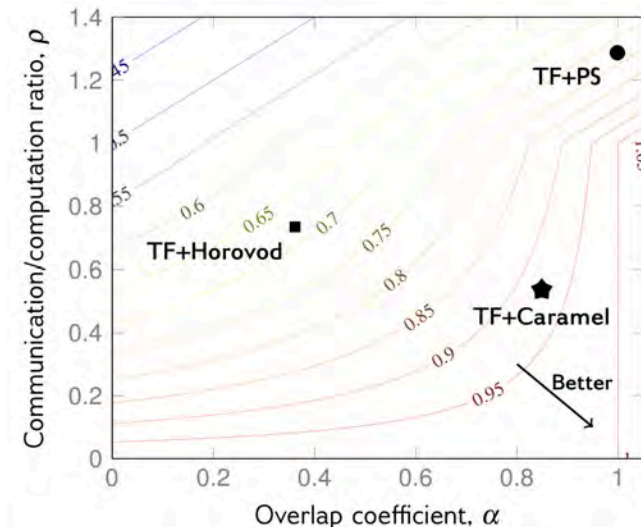
Research Challenge

This work addresses sustained performance and scalability of distributed Deep Neural Networks (DNN) in which training is bottlenecked by parameter aggregation among participating nodes. Two high-level causes of the inefficiency are the communcation patterns of parameter aggregation and the scheduling of operations that can stall available GPU resources when parameters are unavailable. In this work, our goal was to minimize network bottlenecks in distributed DNN training to reduce the iteration time and increase GPU utilization.

Methods & Codes

- Analyzed all-to-one and decentralized data aggregation techniques such as the *bucket* and *halving–doubling* (HD) algorithms
- Investigated dataflow models associated with 16 DNNs and identified common model characteristics that enable efficient network transfers
- Developed Caramel, a model-aware approach to take advantage of faster networks while achieving the highest performance

Results & Impact

- TensorFlow with Parameter Server offers high communication overlap, but with high communications cost.
- The Horovod implementation reduces communication cost with an efficient aggregation pattern but its decentralized patterns suffer from a poor overlap of communication and computation.
- Caramel can improve the iteration time by up to 3.84 times (in VGG-16) and GPU utilization by up to 2.46 times (in AlexNet-v2)

Why Blue Waters

The BlueWaters platform makes it easy to conduct large-scale exploration to find potential performance opportunities. Furthermore, the vibrant community of Blue Waters users and staff helped us to get up to speed faster by leveraging their knowledge and experience.

Allocation: Exploratory/50 knh

Seid Koric

University of Illinois at Urbana-Champaign

Physics & Engineering

PETASCALE MULTIPHYSICS MODELING
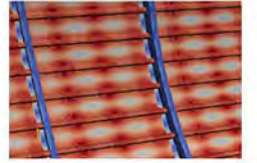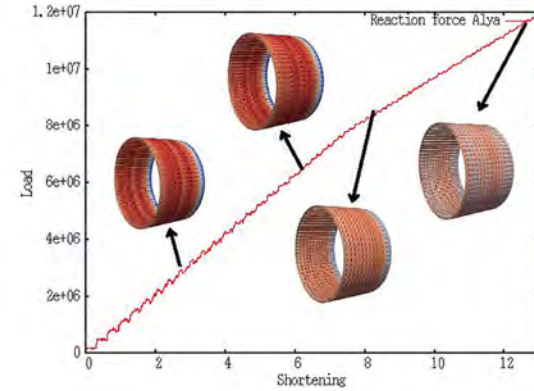
Fuselage barrel 4M mesh and results.

Research Challenge

Provide massively parallel multiphysics finite element code Alya with a fast and efficient direct solver to deal with solid mechanics simulations in which using an iterative solver together with preconditioning techniques is either too expensive or it does not converge to the desired solution. This enables structural analysis of thin structures such as fuselages to determine a panel's buckling and postbuckling capacity, toward assuring the structural integrity up to the ultimate load.

Methods & Codes

- Alya, the multiphysics code developed at the Computer Applications in Science and Engineering (CASE) department of the Barcelona Supercomputing Center
- The WSMP solver integrated into Alya's workflow for symbolic and numerical factorizations as well as solution steps

Results & Impact

The team validated the WSMP implementation in Alya while solving a challenging large ill-conditioned structural problem of great practical importance for society inside the Alya's Solidz module. This was a highly dynamic and nonlinear buckling and postbuckling analysis of a full fuselage barrel loaded in uniaxial compression (see Figure) as validated within the European project SHERLOC (CS2-AIR-GAM-2014-2015-01).

Why Blue Waters

Blue Waters is the only resource where massively parallel multiphysics codes such as Alya with the WSMP solver can be tested by taking full advantage of large amounts of distributed memory, hundreds of thousands of computing cores, and the low latencies and increased bandwidth of leading interconnect network technologies.

CINEMATIC SCIENTIFIC DATA VISUALIZATION FOR CADENS

Visualizing the first 24 hours after the collision that formed the moon, as a Mars-sized body strikes the early Earth. (To appear in *Birth of Planet Earth*.)

Research Challenge

The Advanced Visualization Lab (AVL) team's challenge for this project was to draw on data from scientists involved in high-performance computing-based research to produce and render cinema-quality visualizations for the upcoming full-dome planetarium show, *Birth of Planet Earth*. The effort is part of AVL's continuing work on the NSF-funded CADENS (Centrality of Advanced Digitally Enabled Science) project. The film will employ these visualizations to explore important questions about Earth's formation, and what that history tells us about the chances of finding other planets capable of sustaining life.

Methods & Codes

The team's work depended on several externally provided software packages: Houdini, a visual effects software package, yt for ingesting and re-gridding data, and VMD (Visual Molecular Dynamics) from John Stone at the Biophysics group at the University of Illinois Urbana-Champaign. AVL also developed its own software tools, including Ytini for yt-Houdini integration, and Bluerend to organize the rendering workflow.

Results & Impact

Blue Waters allowed the team to create a refine two data-driven cinematic animations; both will appear in the full-dome planetarium show, *Birth of Planet Earth*, to be released in 2019:

- Formation of the Moon – the first 24 hours after the collision that formed Earth's moon
- Visualizing Energy Harvesting in a Photosynthetic Purple Bacterium

A work-in-progress trailer for the film is available at https://vimeo.com/277190989

Why Blue Waters

Blue Waters allowed the team to iterate quickly and meet deadlines. Over the course of a single weekend, the group rendered a complex scene made up of six separate layers and 20,870 image frames. That ability to render this large set of images in a short period of time allowed AVL to make several iterations before finalizing a video to send to the International Planetarium Society 2018 conference in Toulouse, France.

DEEP LEARNING APPLICATIONS IN ENGINEERING AND QUANTITATIVE FINANCE

Research Challenge

The research team used the Blue Waters supercomputer to support research into applications of deep learning in engineering and quantitative finance. The research challenge is to develop a deep learning method for solving high-dimensional partial differential equations as well as develop a deep learning model for high-frequency financial data.

Methods & Codes

- Machine learning, deep neural networks
- PyTorch, Tensorflow
- GPUs

Results & Impact

Due to the success of deep learning in traditional computer science fields, there is now significant interest in applying deep learning techniques in engineering, science, medicine, and finance. The team analyzed a large high-frequency dataset of electronic market quotes and transactions for U.S. equities and found several interesting insights into the relationship between price formation and order flow.

Why Blue Waters

Deep learning uses multilayer neural networks (i.e., deep neural networks) to build statistical models of data. This training of the deep learning model can be computationally intensive due to both the large number of parameters and the large amounts of data. The team leveraged Blue Waters' large amount of graphics processing units (GPUs) to develop deep learning models for applications in engineering and finance.

PARALLEL ALGORITHMS FOR SOLVING LARGE MULTIDIMENSIONAL ASSIGNMENT PROBLEMS WITH DECOMPOSABLE COSTS

Speedup factors for MDADC (Multidimensional Assignment Problem with Decomposable Costs) instance with $N = 200$ and $K = 200$

Research Challenge

The objective of this research is to develop fast and scalable GPU-based parallel algorithms that can solve large instances of the Multidimensional Assignment Problem with Decomposable Costs which is a fundamental problem in data science applications involving multiple data sources. Data gathered by these sources may be in different formats, and the goal of data association is to merge these data into a cumulative evidence aggregate that can be used for sense-making tasks or to obtain information about the current state of the real world

Methods & Codes

- MPI
- CUDA
- GPU-accelerated Lagrangian heuristic

Results & Impact

Single-GPU experiments revealed that the proposed algorithm is over 200 times faster than the Map-Reduce implementation in previously published work for some of the smaller problem instances.

Why Blue Waters

GPU-accelerated Lagrangian heuristic benefits from the host of powerful GPU-enabled processors available on Blue Waters. The team used over 128 XK compute nodes, which would have incurred significant costs on the proprietary systems such as the AWS.

Allocation: Innovation & Exploration/10 knh

Indranil Gupta

University of Illinois at Urbana-Champaign

Computer Science & Engineering

TOPOLOGY-AWARE DISTRIBUTED GRAPH PROCESSING FOR TIGHTLY COUPLED CLUSTERS

Research Challenge

Massive graphs from social networks provide useful insights about the nature of interconnected systems. Investigated how one can run cloud applications in tightly coupled clusters and network topologies on supercomputers. The key to achieving performance in distributed graph processing systems is to partition the graph intelligently across the nodes of the cluster. This research presented new partitioning techniques that are topology-aware to avoid sending data over long routes.

Methods & Codes

- 'Grid Centroid' partitioning method determines the nodes that should perform state computations for each vertex to minimize network communication cost with other mirrors.

- 'Restricted Oblivious' partitioning method uses a greedy heuristic while creating vertex mirrors to ensure they are all placed close to each other and to the node performing computation.

Results & Impact

Compared to previously existing work, these new partitioning methods produce 25% ~ 33% improvement in runtime, along with a sizable reduction in network traffic (number of hops) in the supercomputer topology.

Why Blue Waters

Blue Waters' 3D Torus interconnect provides a topology with heterogeneous communication costs. It was thus an important part of our experiments.

RADICAL-CYBERTOOLS: A BUILDING-BLOCK APPROACH TO HARNESSING THE POWER OF MANY
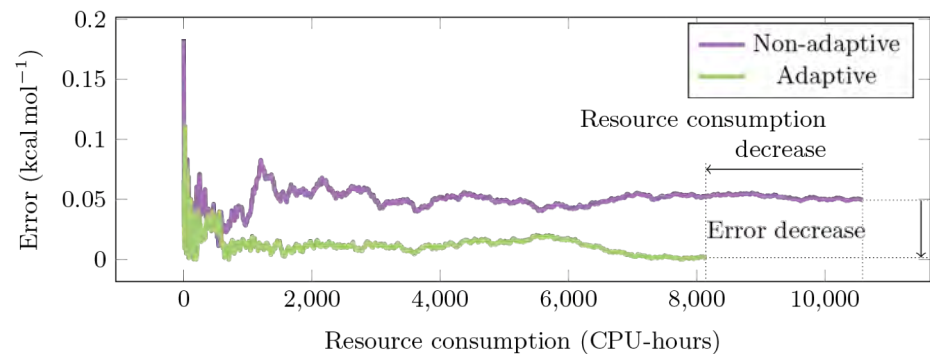
Research Challenge

RADICAL-Cybertools has been used to develop toolkits which have been utilized to support scalable science applications: HTBAC (High-Throughput Binding Affinity Calculator)—HTBAC aims to facilitate the running of drug-screening campaigns of molecular simulations for computing binding affinities. ICEBERG(Imagery Cyberinfrastructure and Extensible Building Blocks to Enhance Research in the Geosciences)—ICEBERG aims to facilitate pan-Arctic and pan-Antarctic science projects requiring manipulation and interpretation of high-resolution satellite imagery.

Adaptive simulation methodologies reduce both resource consumption and uncertainty errors in free energy, leading to a reduction in time-to-solution and improvement in accuracy. The "optimal" simulation parameters are difficult to determine *a priori*, thus requiring runtime determination and adaptation of the workflow based on intermediately generated data.

Methods & Codes

The HTBAC toolkit addresses the need for scalable and accurate protein–drug binding affinity results in acute precision medicine. ICEBERG will provide deep-learning workflows for classification of high-resolution satellite imagery, and its integration with other data streams (e.g. Digital Elevation Models).

Results & Impact

The team was recently awarded the 2018 IEEE SCALE Challenge for using HTBAC on Blue Waters to explore simulations that calculate the strength of drug binding in order to reduce costs and improve computational efficiency for drug design and to underpin development in personalized medicine.

Why Blue Waters

The system is essential for the computational campaign to reach the scales needed to make a clinical impact. HTBAC requires a scalability to at least 2,000 nodes per use-case. The full use-case of ICEBERG requires scalability up to 1,000 GPU nodes to classify images using deep neural networks.

Biology, Chemistry & Health

OUTPERFORMING NATURE: SYNTHETIC ENZYME BUILT FROM DNA FLIPS LIPIDS OF BIOLOGICAL MEMBRANES AT RECORD RATES
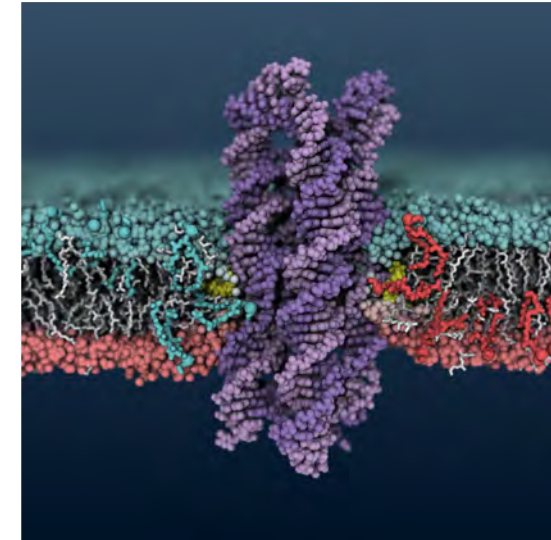
Lipid scrambling by a designer DNA enzyme. The DNA nanostructure (purple) is inserted via hydrophobic cholesterol anchors (yellow) into a lipid bilayer membrane (cut away for clarity). A toroidal pore that connects the upper (blue) and lower (red) leaflets of the bilayer allows lipids to pass between the leaflets.

Research Challenge

Mimicking enzyme function and increasing the performance of naturally evolved proteins is one of the most challenging and intriguing aims of nanoscience. The team employs DNA nanotechnology to design a synthetic enzyme to substantially outperform the biological archetypes. Since the function of an enzyme sensitively depends on its atomic-scale structure and dynamics, this aspect of the studied system necessitates the use of all-atom molecular dynamics (MD) for prototyping the designer enzymes.

Methods & Codes

This project employs the NAMD package to perform explicit-solvent all-atom MD simulations of a synthetic scramblase—an enzyme that transports lipids from one leaflet of a bilayer to the other. The studied system consists of eight DNA strands embedded in a lipid bilayer membrane through two covalently attached cholesterol anchors. Experimental work conducted in the Keyser Lab in Cambridge, UK complements the microsecond-timescale MD simulations.

Results & Impact

The conducted MD simulations provide a microscopic detail of the mechanism of lipid transport, and show that a membrane-spanning single DNA helix forms a toroidal pore that provides a pathway for lipid molecules to cross from one leaflet of the bilayer to the other leaflet. Computations predict very small ~2 k_BT barrier to lipid crossing at the toroidal pore indicating the diffusion-limited rate. The computed rate that exceeds 10^7 molecules per second matches the experiment, which confirms the enzyme activity in cancer cells.

Why Blue Waters

Accurate characterization of structural fluctuations and transport properties of DNA nanotechnology systems enhanced by the presence of nonstandard functional groups requires the use of explicit-solvent all-atom MD simulations. The massive size of DNA structures makes their MD simulations computationally demanding. The large number of GPU-accelerated nodes and fast Gemini interconnect available on Blue Waters significantly speed up the DNA nanotechnology simulations.
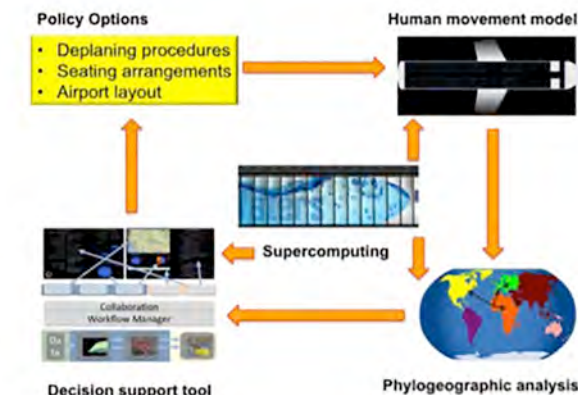
Allocation: NSF PRAC/350 Knh

PI: Ashok Srinivasan

Florida State University

Biology, Chemistry & Health

SIMULATION OF VIRAL INFECTION PROPAGATION THROUGH AIR TRAVEL

Workflow linking fine-scaled pedestrian dynamics model with infection transmission models

Research Challenge

The goal is to develop models and a novel methodology that can provide insight to decision-makers on policies and procedures that will reduce the likelihood of infection spread during air travel. In addition, our research contributions promise major advances in the disciplinary areas of our expertise—pedestrian movement modeling, mathematics, epidemic modeling, computer science, and bioinformatics—with a consequent transformative effect on transportation infrastructure and management. Inherent uncertainties in human behavior and insufficient data during the initial stages of an epidemic make prediction difficult.

Methods & Codes

Pedestrian movement (trajectory and speed) during air travel is modeled as particles based on a force-field approach, and is then integrated with a discrete-time stochastic Susceptible-Infected model for infection transmission that accounts for demographic stochasticity and variations in the susceptibility of the population. This information is input to a global phylogeography model, implemented using the BEAST software package, that parameterizes the sources of uncertainty and evaluates vulnerability under different possible scenarios.

Results & Impact

Factors such as random (*i.e.*, not zone-based) boarding, in-plane movement, deplaning procedure, seating arrangement, plane sizes, and using temporary walls instead of ropes for queue management can lead to substantial reduction in infection transmission over the course of an epidemic. New research into other directly transmitted diseases requires changes in our approach that include aerosol and fomite transmission mechanisms, while pedestrian movement accounts for proximity among infectious and susceptible individuals.

Why Blue Waters

In a new emergency, one usually needs to model a variety of scenarios, leading to a large parameter space of uncertainties. This requires a large computational effort involving a low-discrepancy parameter sweep and iterative fine-tuning. Rapid turn-around time during the course of a decision meeting is critical, and the resulting computational load requires massive parallelism only available on Blue Waters. We obtained support from the Blue Waters team to optimize parallel IO in our code, reducing simulation time by a factor of two.

MODELING AND SIMULATIONS OF COMPLEX DYNAMIC MUSCULOSKELETAL ARCHITECTURES

Top view of three thumb-sized biobots in a culture dish. Modeling and simulations were employed to design and optimize function and performance. These biohybrid machines walk on the dish surface at about one millimeter per second.

Research Challenge

Bioinspired approaches rely on mimicking existing natural solutions. However, natural creatures have not evolved to optimize engineering objectives. There exists limited information to determine whether a particular solution is optimal. Moreover, today's materials and components pose constraints and enable opportunities that may differ from their biological counterparts. This team suggests that manmade solutions obtained through inverse design based on an automated optimization process may outperform pure biomimicry.

Methods & Codes

This project has been developing and implementing novel schemes for the direct numerical simulation of swimming bodies. Their algorithms rely on remeshed vortex methods enhanced with projection approaches to capture the effects of the fluid on the body and, with penalization techniques, that of the body on the fluid. This methodology is coupled with a musculoskeletal solver that captures the compliant dynamics of musculoskeletal systems made of bones, tendons, and muscles.

Results & Impact

- This team has developed *Elastica*, an application able to capture the dynamic response of complex musculoskeletal structures.
- *Elastica* was used to computationally design, simulate, and optimize for the first time the structure of a biohybrid walking bot. In collaboration with experimentalists at the Micro and Nanotechnology Laboratory at Illinois, this design was fabricated and tested, leading to the largest, fastest locomotive bot to date

Why Blue Waters

Blue Waters' sheer size and cutting-edge technology enable optimization processes that entail thousands of simulations. This allows the design of unprecedented biological architectures, bringing within reach novel high-impact applications, from soft robotics and biomedicine to precision manipulation and fabrication.

Allocation: BW Professor/200 Knh
PI: So Hirata
University of Illinois at Urbana-Champaign
Biology, Chemistry & Health

TOWARD PREDICTIVE COMPUTATIONAL DESIGN OF PRECISION MOLECULAR OPTOELECTRONICS
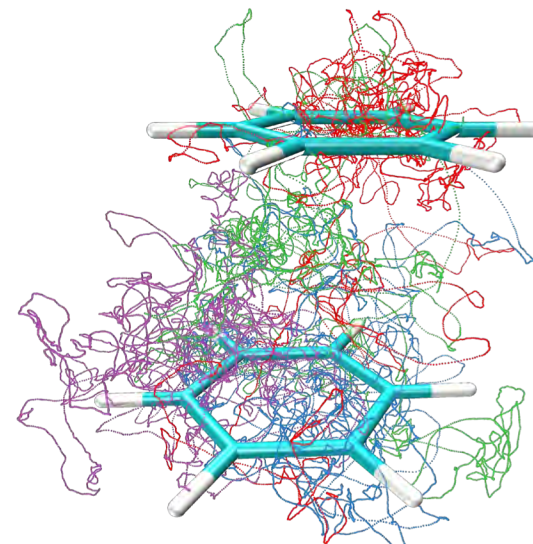
Research Challenge

Chemical technology can fabricate solid-state materials made of organic components with precise dimensions for optoelectronic device and catalytic applications. An even greater advance occurs when this synthetic capability is coupled with computational machinery that can predict their properties and functions quantitatively. *Ab initio* many-electron theory provides the necessary foundation, but its legacy algorithms based on matrix algebra are poorly scalable for larger molecules or larger computers. This project introduces completely new and scalable stochastic algorithms.

A correlated random walk of an electron pair in imaginary time, efficiently solving the fundamental equation of motion of quantum chemistry numerically.

Methods & Codes

The usual sum-of-products matrix expressions of second-order MBGF (GF2) theory and its complete-basis-set (CBS) correction by explicitly correlated (F12) ansätze are mathematically transformed into high-dimensional integrals, which are then evaluated by a highly scalable Metropolis Monte Carlo (MC) method. The resulting stochastic methods—MC-GF2 and MC-GF2-F12—can compute energy differences directly without a sign problem in a scalable manner with respect to both compute size and system size.

Results & Impact

The developed MC-GF2-F12 method efficiently executes on both XE and XK nodes. It enables an exact (CBS-limit) GF2 calculation of electron-detachment energies for a wide range of large conjugated organic molecules. The largest calculation for C_{70} with 1,610 basis functions was run on 128 GPUs for the GF2 portion and on 896 CPUs for the F12 part. The implemented two-level parallelism enhances the performance of the redundant-walker algorithm on GPUs beyond the degree that is possible by merely running it on many CPUs.

Why Blue Waters

The stability and ease of use of Blue Waters as well as the balanced deployment of CPUs and GPUs are all essential for rapid coding/profiling of new scalable algorithms from scratch and their capacity testing.

Allocation: Director Discretionary/50 knh

PI: Mohammed El-Kebir

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

ALGORITHM FOR RECONSTRUCTING THE LIFE HISTORY OF TUMORS FROM GENOMICS DATA
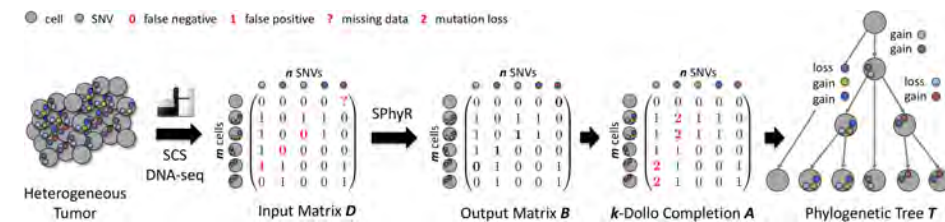
Research Challenge

The phylogeny estimation problem from single-cell sequencing data is a variant of the classic phylogeny estimation problem with incorrect and missing data due to the used sequencing technology. Current methods aim to simultaneously construct a phylogenetic tree and correct these measurement errors using either too stringent or too permissive evolutionary models. There is a need for methods that employ appropriate evolutionary models that strike a balance between being realistic and yet, sufficiently constrained.

Tumors are composed of cellular populations with distinct sets of mutations. Single-cell sequencing of a tumor yields an input matrix D with incorrect and missing entries. SPhyR aims to simultaneously correct errors in matrix D and infer the evolutionary history of the cells.

Methods & Codes

To solve the phylogeny estimation problem from single-cell sequencing data, the researchers used techniques from combinatorial optimization. More specifically, they formulated the problem as an integer linear program (ILP) and designed and implemented a custom column generation and cutting plane approach for the ILP formulation.

Results & Impact

In light of frequent loss of point mutations in cancer due to copy-number aberrations, the k-Dollo model is more appropriate than the evolutionary models utilized by previous methods. This project resulted in a novel combinatorial characterization of solutions to the underlying computational problem as constrained integer matrix completions, which formed the basis for the efficient integer linear programming approach utilized by SPhyR.

Why Blue Waters

The computational resources of Blue Waters allowed the research team to perform these experiments at scale, enabling them to study the performance of their algorithms and the underlying problem statements in many different experimental settings.

Allocation: Exploratory/50 knh

Andrew Ferguson

University of Chicago

Biology, Chemistry & Health

IN SILICO VACCINE DESIGN THROUGH EMPIRICAL FITNESS LANDSCAPES AND POPULATION DYNAMICS

The black crosses characterize the immune response of the inoculated host relative to no vaccination (green circle). Red circles indicate the optimal vaccine candidates. Blue circles indicate previously identified vaccine candidate. Candidates in the upper-left quadrant provide superior strength and length of control relative to no vaccination.
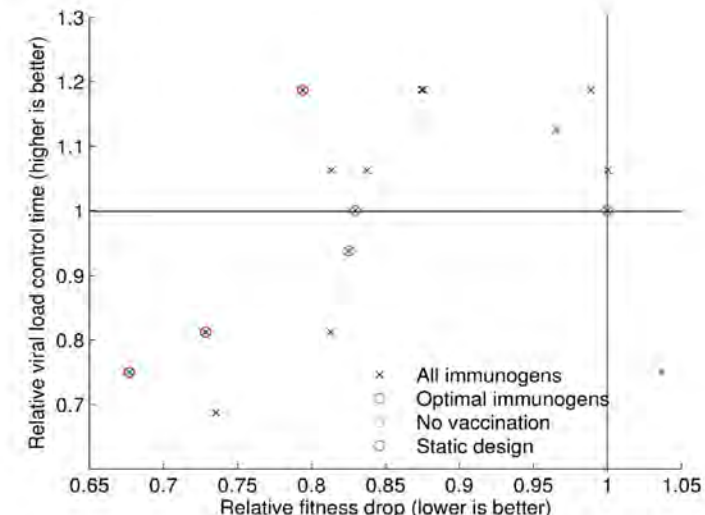
Research Challenge

The hepatitis C virus (HCV) poses a global threat to public health. A prophylactic vaccine represents the most cost-effective and realistic strategy to combat the epidemic, but despite 25 years of research, a vaccine is still not available. Computational models of viral infection and the host immune response can systematically identify promising targets that may be translated into rational precepts for experimental development and testing of HCV vaccines.

Methods & Codes

- An agent-based model comprising 50,000 distinct viral sequences
- Ordinary differential equations describing the host immune response
- Coupling to the viral dynamics via tracking viral strains recognized and attacked by T-cells
- A stochastic integration protocol via Gillespie dynamics to explicitly model the fluctuations that are important when simulating small numbers of cells

Results & Impact

For two representative hosts, the team used their simulator to predict the efficacy of the ensemble of all possible vaccine candidates consistent with the immunological genotypes of the hosts. In each case, they identified a number of promising vaccine candidates that led to strong and durable responses by priming T-cells that imposed strong fitness penalties upon the viral population for long periods of time.

Why Blue Waters

The scale and parallelism available within Blue Waters support the computational intensity of each single simulation and also enable the range of simulations to evaluate large numbers of vaccine candidates in a variety of hosts. Furthermore, the volume of data generated is also significant as the viral sequences present at each time point must be written to disc.
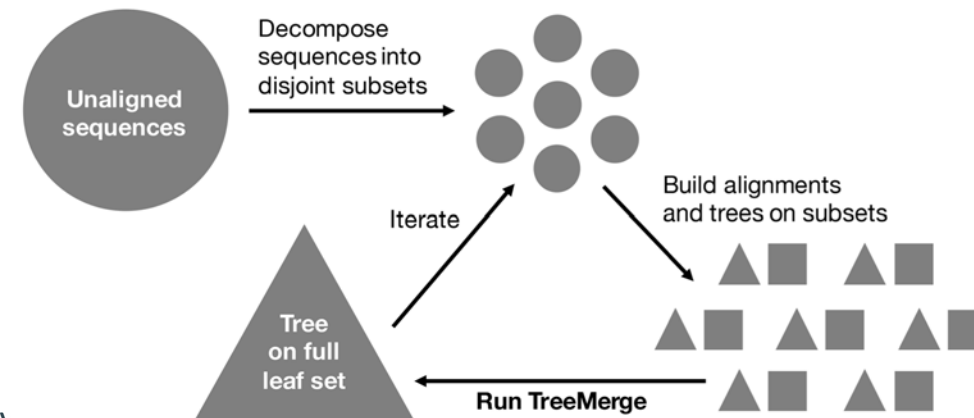
Allocation: Exploratory/50 knh

PI: William Gropp

Co Pi's: Tandy Warnow, Erin Molloy

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

CONSTRUCTING LARGE EVOLUTIONARY TREES ON SUPERCOMPUTERS

Research Challenge

Evolutionary trees (called phylogenies) are important for addressing many questions in basic biology and applied research (for example, phylogenies can be used to identify microorganisms living in the human gut). However, phylogeny estimation from genetic data is computationally challenging. The leading approaches to phylogeny estimation have two steps: 1) a multiple sequence alignment is estimated from the input genetic data and 2) a Maximum Likelihood tree is estimated from the alignment. Both steps do not scale to genetic datasets with large numbers of unaligned sequences.

Figure 1: **TERADACTAL** divides unaligned sequences into **disjoint subsets** (circles), builds alignments (squares) and trees (triangles) on each subset, and then merges the subset trees together in polynomial time using a novel technique, called **TreeMerge**. This process can iterate by decomposing the final tree into subsets.

Methods & Codes

The team's approach, called **TERADACTAL**, is shown in Figure 1. The divide-and-conquer approach enables alignments and trees to be estimated on subsets in parallel; however, the innovation is a newly developed technique, called TreeMerge, that combines subset trees in polynomial time without sacrificing accuracy. Because TreeMerge combines subset trees in a pair-wise fashion, the conquer phase can also be performed in parallel. TERADACTAL is freely available on Github (https://github.com/ekmolloy/teradactal-prototype).

Results & Impact

TERADACTAL achieved similar error rates (within 1-3%) of the best two-phase methods tested (e.g., RAxML given the true alignment). Thus, TERADACTAL can achieve similar error rates to the leading two-phase methods but is highly parallelizable and avoids computationally challenging tasks, such as, alignment estimation on the full dataset, Maximum Likelihood tree estimation on the full dataset, and supertree estimation. **This work is a major advancement toward constructing the Tree of Life using supercomputers.**

Why Blue Waters

Blue Waters was used to demonstrate that existing parallel codes (e.g., PASTA and RAxML) could not effectively run on datasets with one million sequences on Blue Waters, to extensively test the TERADACAL prototype, and to compare the TERADACTAL prototype to the leading two-phase methods. Specifically, the team performed a large simulation study requiring over 36,000 node hours. These analyses were completed in under a month but would have required over a year to run on other available systems.

Growth of the abdominal aortic aneurysm (colored by the deformation of the artery wall).

BLOOD-ARTERY INTERACTION AND SIMULATION-BASED ANALYSIS OF AORTIC ANEURYSM
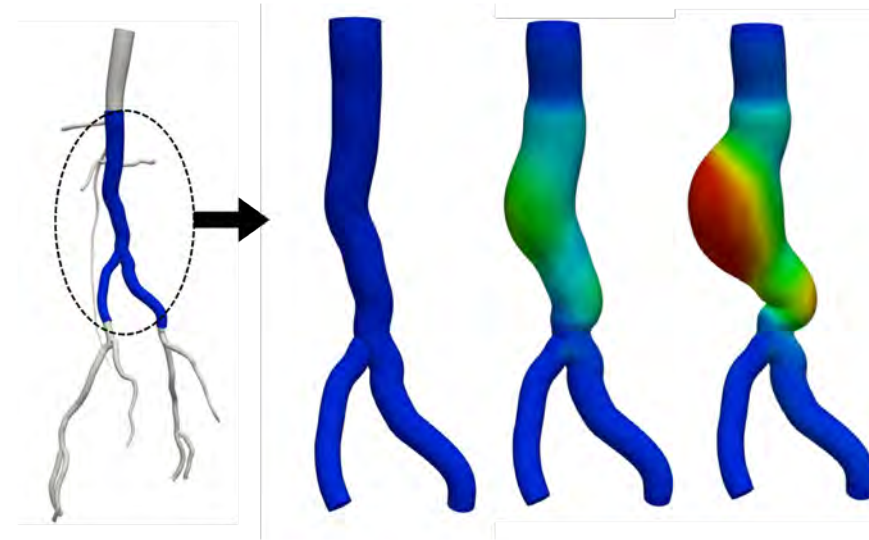
Research Challenge

The research team is developing new finite element method and associated computer code for large deformation modeling of blood–artery interaction and application to the challenging problem of abdominal aortic aneurysm, which is the 10th leading cause of adult death in the United States. These computer codes can be used for modeling the long term progression of the arterial disease as well as local changes in the flow patterns that arise due to marked changes in the geometric structure of the arterial system.

Methods & Codes

- Patient-specific geometric models of the aortic and femoral arteries constructed from CT scans
- Blood modeled as a non-Newtonian fluid, while the deforming artery was modeled as a hyperelastic solid with fiber reinforcement in the circumferential direction
- PETSc

Results & Impact

The team has shown that advanced Fluid–Structure simulations that can provide important data for the diagnosis and treatment of arterial diseases such as an aneurysm are critical in planning and developing strategies to cure these diseases. The team was able to identify high-viscosity regions where blood coagulation can potentially take place and also to project the arterial Wall Shear Stress (WSS) on the arterial walls (WSS is one of the most significant factors for the progression of arterial disease).

Why Blue Waters

The team's algorithms for Computational Fluid–Structure analysis are amenable to efficient parallelization because major portions of the computations are carried out at the local-element level. Smart implementations of the method can take advantage of the local memory that available at the processing nodes. These algorithms are well suited for distributed memory parallelism, and Blue Waters provides an ideal platform for implementation of such algorithms.
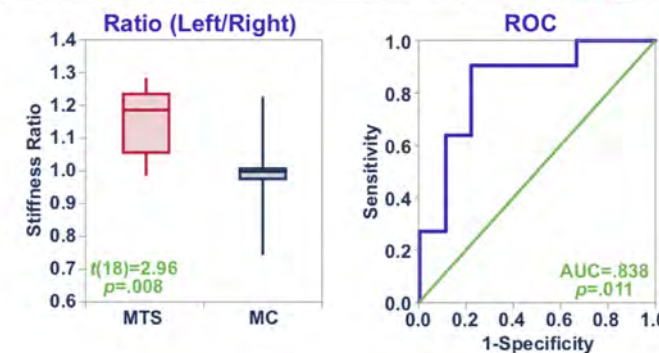
Allocation: Illinois/100 Knh

PI: Brad Sutton

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

HPC-BASED HIGH-RESOLUTION QUANTITATIVE MRI

Research Challenge

The team is a leader in Magnetic Resonance Elastography (MRE) development and application, enabling the study of the mechanical properties of the brain. They have been able to map the brain at unprecedented resolution and to probe specific neuroanatomical structures. The method will benefit investigators of both neurodegenerative disease and epileptologists looking at long-term changes to tissue.

(Top) MPRAGE, FLAIR, and MRE-Stiffness images from a patient with mesial temporal lobe sclerosis. (Bottom) The ratio of stiffness of epileptogenic to nonaffected hippocampus (HC) is significantly greater compared to controls. The ROC curve indicates that the ratio of HC stiffness dissociates patients from controls with high accuracy.

Methods & Codes

PowerGrid, an open source, GPU-accelerated, MRI image reconstruction toolkit, computes of a series of displacement encoded images per subject to enable MRE. PowerGrid leverages GPUs to apply parallel imaging, corrections for imperfect magnetic fields, and subject motion. Nonlinear Inversion (NLI) for iterative MRE material property estimation from displacement encoded images, divides the brain into smaller regions for Finite-Element Method (FEM) meshing and optimization of properties at a local level, which are then reconstituted as a global solution, allowing parallelization across hundreds to thousands of CPU cores.

Results & Impact

MRE is a useful tool for identifying patients with mesial temporal lobe epilepsy due to an asymmetry in hippocampal scarring. The research work has revealed significant softening of the nonsclerotic hippocampus compared to healthy controls, which may provide an important biomarker for early disease detection, which is critical for successful treatment before significant and irreversible damage occurs. The team now has sufficient pilot data to write a grant to translate this technology into the clinic for early detection.

Why Blue Waters

The Blue Waters system provides a unique resource for MRE because it allows for rapid transfer of imaging data from the scanners and the low-latency memory transfers necessary for highly parallel computations. For PowerGrid, Blue Waters provides an environment with a large number of GPUs as well as support for MPI for leveraging GPUs across nodes.

STRUCTURAL BASIS FOR EXTREME COLD TOLERANCE IN THE EYE LENSES OF TELEOST FISHES

Research Challenge

Lenses of teleost fishes have to maintain transparency despite their constituent lens proteins facing substantial pressure to unfavorably interact with each other at very high concentrations and temperatures much colder than that of mammalian lenses, and we suspect they exhibit enhanced flexibility that prevents these interactions from occurring; given the overwhelming computational resources needed to model the biophysical behavior of the lens proteins, we require the resources of Blue Waters to address this hypothesis.

(a) RMSF of eight mammalian γ-crystallin isoforms (blue) and 12 zebrafish isoforms (orange) at 0° C. (b) Proportion of γ-crystallins with RMSF differences among zebrafish γ-crystallins (blue) and mammalian γ-crystallins (orange) that are greater than zero at 0° C.

Methods & Codes

This team ran molecular dynamics (MD) simulations on twelve zebrafish and eight mammalian isoforms at a cold temperature (0° C), and at the normal body temperature (25° C and 37° C respectively). Three replicates of each γ-crystallin isoform were simulated for 50 nanoseconds (ns) in NAMD [7] 2.12 using CHARMM27 force field parameters.

Results & Impact

At 0° C, it is evident that zebrafish γ-crystallins are more flexible than the mammalian isoforms across all sites. A distinct disparity in flexibility is present among zebrafish and mammalian γ-crystallins, and the elevated flexibility of teleost γ-crystallins may account for their remarkably cold-tolerant lenses.

Why Blue Waters

Their work requires simulating three trials of 49 proteins at two temperatures, and over a long timecourse of 50 ns to detect meaningful molecular behavior. Only the petascale computational power and resources of Blue Waters could allow them to achieve this core portion of the project in a reasonable amount of time for downstream analyses to test the hypotheses.

Fig. 1. Conformational change in the strigolactone receptor D14 leading to an open and closed state.

UNRAVELING THE MOLECULAR MAGIC OF WITCHWEED
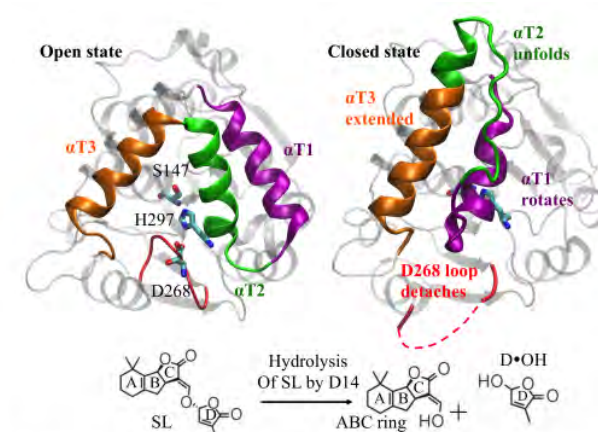
Research Challenge

Witchweed is a root parasite that is considered a serious agricultural pest affecting crops such as sorghum, maize (corn), rice, millet and cowpea, among other crops. Witchweed seeds remain in the soil for decades until favorable germination conditions are provided by a host plant. Globally, it leads to economic loss of $10 billion every year, and loss to livelihood of over 100 million farmers around the world. Currently, there is a need for a control technique to combat the outbreak of this menacing parasite.

Methods & Codes

The goal is to develop a molecular control strategy that inhibits growth of witchweed to avoid crop failure. To do this, there is the need to understand the activation process of the strigolactone receptors to which Witchweed is extremely susceptible to vs the host plants, therefore potentially stopping this pest. The project utilized Molecular Dynamics (MD) simulations that mimic and predict the time evolution of a system of atoms, assuming a given potential energy function.

Results & Impact

The simulations have revealed the flexible regions of the protein involved in the activation process. Also confirmed is the determination that the intermediate-bound protein has more potential activation pathways than the apo protein. The conclusion is that the witchweed protein has lower free energy for activation as compared to Arabidopsis (the host). This might explain the higher SL affinity in witchweed germination.

Why Blue Waters

Understanding the slow conformational transitions in proteins requires hundreds of microsecond-long simulations. Blue Waters provides the state-of-the-art computer architecture needed to perform such studies. This project employs large-scale adaptive sampling protocols, which can be efficiently performed on Blue Waters' GPU and CPU framework. This current work would not be possible without Blue Waters.

Allocation: Illinois/600 Knh

PI: Diwakar Shukla

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

MOLECULAR SIMULATIONS ON BLUE WATERS AID THE UNDERSTANDING OF HOW PLANTS TRANSPORT SUGARS BETWEEN CELLS

MSM-weighted free-energy plots were obtained by projecting the data on extracellular and intracellular gating distances with extended OF state to investigate the glucose recognition, binding, and transport mechanism over a duration of ~68 µs.
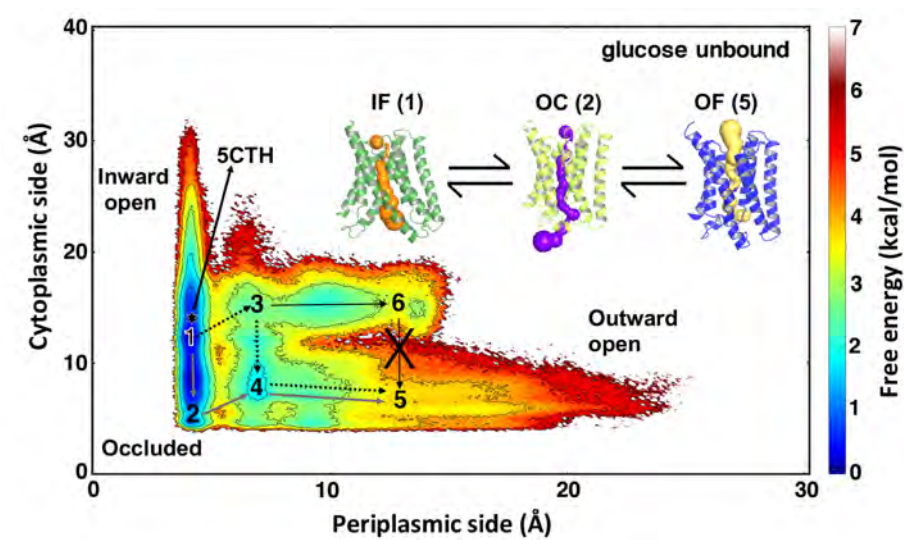
Research Challenge

- An alternative solution to increase crop yield may be to change observable characteristics of an individual plant.

- Experimental studies suggest that some plants store sugar in vascular tissue and transport it from the leaves to various organs.

- The sugars (glucose) that are produced in the leaf during photosynthesis and transported to the phloem via the SWEETs' sugar transporters have the potential to improve crop yeild.

Methods & Codes

- These MD simulations rely on numerical integration of Newton's equation of motion for the interacting atoms to create time-dependent trajectories for all atoms of the system, which together provide a simulation of the biomolecule's dynamical motion.

- Simulations were performed in AMBERv14, analysis was performed with C++ and Python tools, visualization is done with VMD and Pymol.

Results & Impact

- This study provides novel insights into the molecular mechanism of sugar transport in plants helping create a new path to address world hunger.

- This study reveals atomistic-level detail of the conformational dynamics of OsSWEET2b and the free-energy barrier associated with the transport process.

- The results explain how the glucose regulates the SWEET function and key residues involved in the conformational switches.

Why Blue Waters

- This work is computationally demanding and parallel computing reduces the cost and time effectively.

- Blue Waters is a perfect resource that has the massive architecture to run parallel jobs.

- These computations are not possible to perform in a reasonable time without Blue Waters' petascale computing capability.
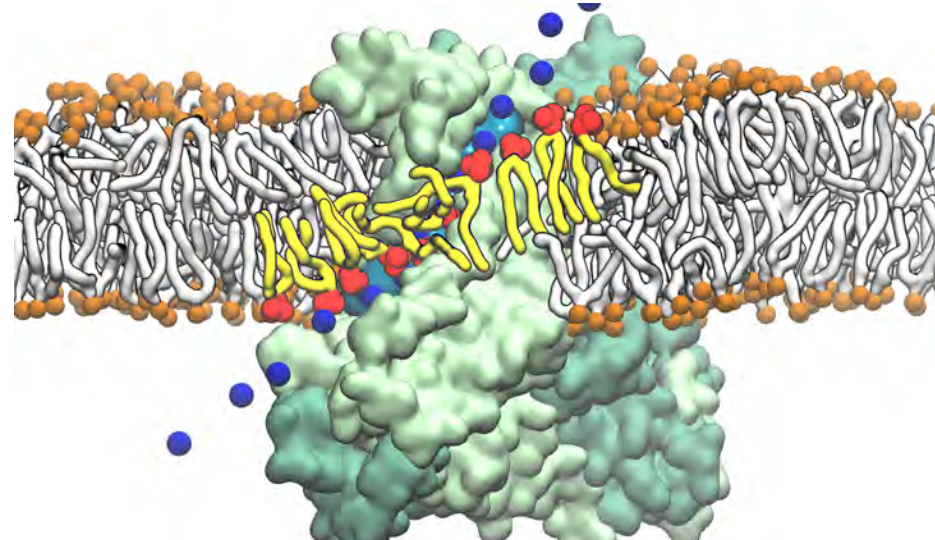
MOLECULAR MECHANISM OF LIPID AND ION TRANSPORT IN A PHOSPHOLIPID SCRAMBLASE

An ion-conducting "proteolipidic" pore formed at the interface of the channel protein TMEM16 (green) and lipid head groups (red) providing a pathway for ions (blue) to cross the membrane.

Research Challenge

Nature distributes different phospholipid species asymmetrically between the two leaflets of the cellular membrane. Dissipation of this asymmetry in response to the elevation of cytoplasmic Ca^{2+} concentration is a ubiquitous signaling mechanism critical for diverse cellular events including blood coagulation, bone mineralization, and cell–cell interaction. Phospholipid scramblases mediate the phospholipid scrambling process. The absence of phospholipids and ionic substrates in their solved structures leaves unanswered the question of how they conducts both lipids and ions.

Methods & Codes

The team conducts extensive MD simulations on the atomic models of the lipid scramblase in asymmetric lipid bilayers in the presence and in absence of Ca^{2+} ions at the activation binding sites. Application of multiple levels of transmembrane voltage to the equilibrated Ca^{2+}-activated structure allows to examine the ion permeation properties and to compute the ionic conductivity across the membrane. All MD simulations employ the NAMD package.

Results & Impact

Simulations explain that thinning and deformation of the lipid bilayer drives lipid translocation. The team detected one spontaneous full lipid scrambling event through the membrane-spanning lipid translocation track under equilibrium conditions, and four full scrambling events in the presence of voltage. Simulation determined key amino acids that enhance scrambling. Experimentally engineered scramblase activity in a homologous Ca^{2+}-activated ion channel confirmed that prediction.

Why Blue Waters

The high-performance architecture of Blue Waters makes it an excellent computing resource for the present study. The project employs the GPU-accelerated simulation program NAMD that has been extensively tested and optimized for Blue Waters. The large number of GPU-accelerated XK nodes significantly increases our computational productivity. Finally, the technical support provided by the Blue Waters team greatly facilitates the achievement of research goals.

Allocation: Illinois/200 Knh

PI: Gustavo Caetano-Anolles

University of Illinois at Urbana-Champaign

Biology, Chemistry and Health

UNTANGLING THE ORIGINS OF PROTEIN FLEXIBILITY AND BIOLOGICAL FUNCTIONS

Research Challenge

Protein loops are flexible elements of macromolecular structures and are promising candidates for uncovering the evolutionary relationship between protein function and dynamics. Employing dynamic networks to summarize the atomic trajectories of loops of metabolic enzymes obtained from molecular dynamics (MD) simulations, the team studies how network structure changes with molecular functions and protein history.

The diameter (y-axis) of dynamic networks across the evolutionary timeline (x-axis) of structural domains spanning 3.8 billion years of evolution. Evolutionary age ranges from the origin of proteins (0) to the present (1). The mean of distributions suggests network size is maintained during protein evolution.

Methods & Codes

The team used the NAMD program to perform 50-70 nanosecond-long molecular dynamics simulations for 116 protein loops. To shortlist proteomes belonging to Archaea, Bacteria, Eukarya, and viruses, the team used the RefSeq database and scanned nearly 2,100 proteomes with protein domain HMM profiles using the HMMER tool. To analyze the MD simulations of protein loops in single-domain meta-consensus enzymes, the team generated dynamic networks of positive and negative correlations of motions based on these simulations.

Results & Impact

The team found that average diameter and length of dynamic networks along the evolutionary timeline of structural domains remains constant throughout the entire 3.8 billion years of evolution. This finding suggests that dynamics is preponderantly a physical property of proteins. Careful analysis of results obtained on Blue Waters suggests that robustness in the structure of dynamic networks is tempered by the emergence of modules of dynamic behavior in specific functions of the molecular systems.
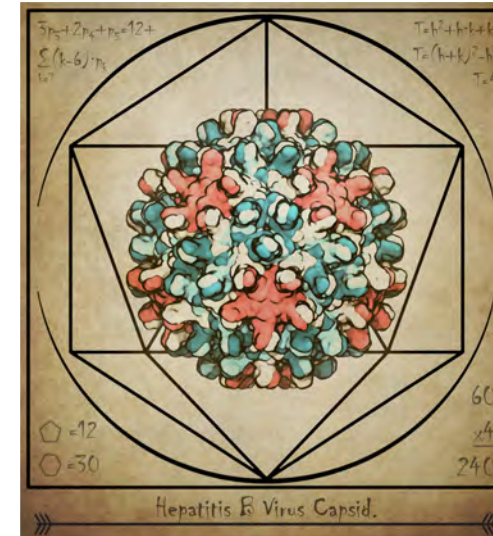
Why Blue Waters

Without access to Blue Waters, this computationally intensive study would not have been possible to achieve in reasonable time. The team commends the Blue Waters support staff – they have been extremely helpful with prompt resolution of computational as well as logistical issues during the execution of this project.

Allocation: Illinois/582 Knh

PI: Juan Perilla, Co-PI: **Jodi Hadden**

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

UNVEILING THE FUNCTIONS OF VIRUS CAPSIDS THROUGH THE COMPUTATIONAL MICROSCOPE

The HBV capsid is an icosahedral protein shell composed of 240 identical proteins. Although experimentally determined structures suggest that the capsid is rigid and symmetric, our simulations on Blue Waters reveal that the capsid is highly flexible and can distort asymmetrically.

Research Challenge

Viral pathogens are a major risk to **public health,** and millions of people die annually due to a lack of effective antiviral treatments. The development of novel drug compounds that can target viruses depends heavily on characterizing the components of virus structure and the roles these components play in facilitating infection. This project specifically studied the viruses that **afflict humans** with **hepatitis B liver disease** and **T-cell leukemia**.

Methods & Codes

Molecular dynamics simulations provide a powerful technique to investigate the dynamical structure and chemical–physical properties of virus capsids. This work utilizes the **NAMD** code to perform **all-atom simulations** of **virus capsids**. Our work has demonstrated that, when performed at the **all-atom level** of detail, simulations are capable of capturing even **subtle effects** on **capsid structure** and dynamics induced by bound **drug** molecules.

Results & Impact

This work studied **Hepatitis B virus** (HBV), a leading cause **of liver disease** worldwide, including cancer, and **Human T-cell leukemia virus type 1**, which causes **cancer**, and is a cousin of the HIV-1 virus, which causes AIDS. Both simulations have provided significant **new insights** into the **capsid structure** suggesting new ways for **drugs to target these viruses**.

Why Blue Waters

Capsid systems encompass millions of atoms, and computing the interactions among such large numbers of particles over microsecond timescales can take months, even on thousands of processors. Further, analysis of the colossal data sets generated by capsid simulations is enabled through access to the massively parallel computing power and high-performance Lustre filesystem provided by Blue Waters. **Importantly, our discoveries were inaccessible to state-of-the-art experimental methods and were made possible only through access to the petascale computing power of Blue Waters.**

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

Allocation: Illinois/350 Knh
PI: Liudmila Mainzer
University of Illinois at Urbana-Champaign
Biology, Chemistry & Health

IDENTIFICATION OF MISSING VARIANTS IN ALZHEIMER'S DISEASE, AND THE NEW STANDARDS FOR GENOMIC VARIANT IDENTIFICATION IN LARGE COHORTS
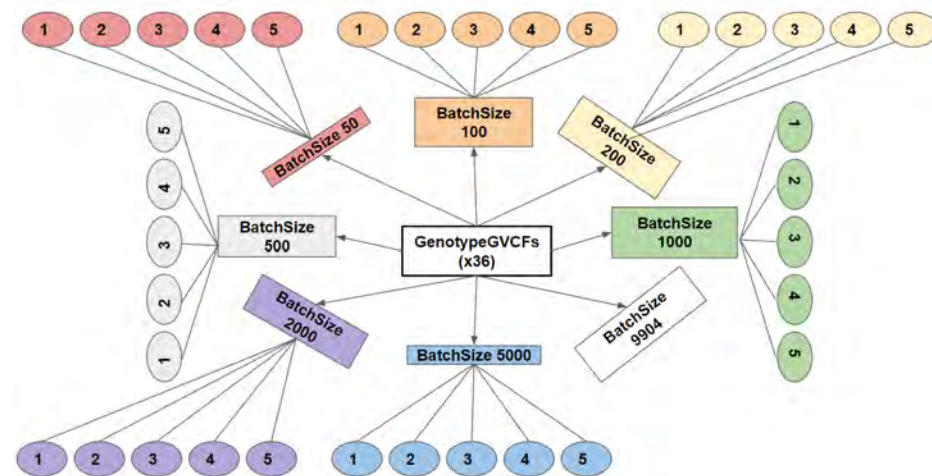
Research Challenge

Alzheimer's disease (AD) is a neurodegenerative dementia that affects more than five million Americans and more than 35 million people worldwide. This project will deliver novel genomic variants that have remained hitherto undetected by the standard workflow in AD. These variants will be posted into public databases for use by researchers and clinicians worldwide to improve understanding of the genomic underpinnings of AD, as well as drug development and treatment outcome prediction.

To study the effects of sample size, we performed joint genotyping on variable-size batches subsampled from 10,000 whole exomes. We subsampled each batch size five times, and ran joint genotyping to see how many variants could be recovered. See https://github.com/dpwickland/JointGenotyping_BW.

Methods & Codes

The team tested a select number of steps and parameters in the variant detection pipeline in the context of sample sizes. They grouped the Alzheimer Disease Sequencing Project (ADSP) samples into different samples sizes of 50, 500, 1000, 2000, 5000, and 10,000. For each sample size, they tested two different aligners, three different variant callers, multi- vs. single-sample variant calling, and five different parameter settings in the variant calling and quality control process. The goal was to build a new set of guidelines for variant discovery based on project size.

Results & Impact

By combining two read aligners and several variant callers into the workflow, they were able to recover 50% of the variants in the ADSP data that were missed by the standard protocol. They further annotated SNPs, or genetic variations in a single DNA building block, as synonymous or nonsynonymous, and assessed the proportion of alternate alleles between cases and controls. They discovered SNPs within genes previously reported to interact with Alzheimer's-related proteins or to function in the brain.

Why Blue Waters

The team tested multiple combinations of steps and parameters using data from the Alzheimer's Disease Sequencing Project consisting of over 10,000 whole exome sequencing samples. The total amount of time that would be required to complete this project on a single server would be 109 years. On Blue Waters, they were able to run a single workflow on the entire set of 10,000 AD samples by parallelizing across thousands of nodes.
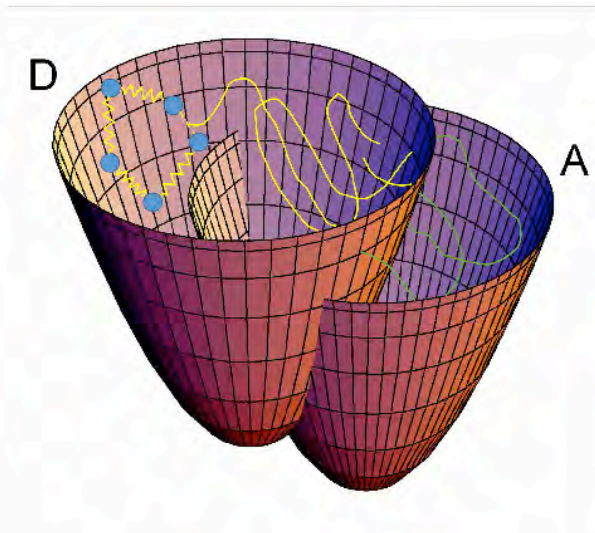
QUANTUM-CLASSICAL PATH INTEGRAL SIMULATION OF ELECTRONIC TRANSITIONS

Research Challenge

Classical molecular dynamics methods are inadequate for describing charge transfer and, more generally, electronic transitions. On the other hand, the cost of quantum mechanical calculations scales exponentially with the number of interacting particles. We have introduced a rigorous quantum–classical path integral (QCPI) formulation, which treats the interaction between quantum and classical degrees of freedom in full atomistic detail without assumptions. However, the QCPI expression involves an astronomical number of terms and thus appears impractical.

Schematic illustration of a classical trajectory with initial conditions sampled from a path integral-quantized phase space density and hopping between the donor and acceptor potential surfaces corresponding to the two charge states of an electron transfer pair according to instantaneous quantum states.

Methods & Codes

Several advances in the understanding of interference and decoherence have made the QCPI methodology practical for the simulation of condensed-phase reactive processes. Current work implements a new acceleration of the code, which is based on a decomposition of the path integral that leads to a dramatic reduction of CPU time. Further, the QCPI code is augmented to allow the treatment of zero-point energy in the classical degrees of freedom.

Results & Impact

The QCPI methodology enables the simulation of charge-transfer reactions in solution with unprecedented accuracy. The ability to perform all-atom calculations with potential interactions treated in full detail leads to results of unparalleled precision. These calculations shed light on the complex interplay among molecular/solvent timescales, electronic couplings, and reorganization energy.

Why Blue Waters

The QCPI formulation is well suited to a decomposition based on multilevel parallelism, and Blue Waters provides the ideal platform for its implementation. Moreover, because the trajectories are independent and generally relatively short, it is possible to assign a single trajectory to each core within a given processor while maintaining computational efficiency. This multilevel approach has the benefit of minimizing communication time while maximizing concurrent processing .
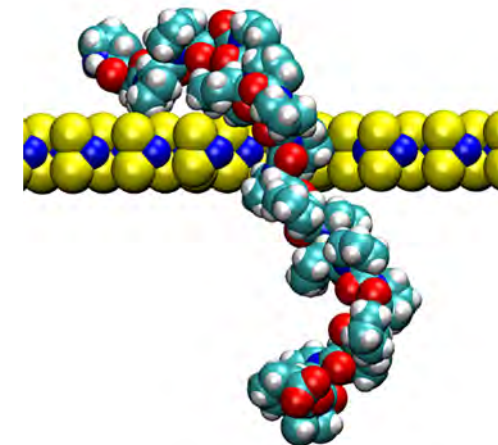
Allocation: Illinois/135 Knh
PI: Narayana Aluru
University of Illinois at Urbana-Champaign
Materials, Biology, Chemistry and Health

DETECTION OF AMINO ACIDS WITH HIGHLY SENSITIVE MoS2 NANOPORES: TOWARD MACHINE LEARNING-BASED PREDICTIVE MODELS

A snapshot of proline polypeptide translocation through the MoS2 nanopore.

Research Challenge

Designing biological and synthetic nanopores with predefined properties for molecular transport is one of the most challenging problems in biotechnology. Biological and synthetic nanopores have been employed for label-free, high-resolution sequencing of DNA. The challenges associated with identification of biological molecules using nanopores are the low signal-to-noise ratio, pore degradation, unique identification of individual molecular units in real time, and the high speed of molecular movement through a nanopore

Methods & Codes

We performed Quantum Calculations based on VASP to determine interaction potentials and molecular dynamics simulations using LAMMPS to determine trajectories of amino acids. Each simulation box comprised of about 32,000 atoms contained a monolayer membrane of MoS_2, an amino acid chain, water molecules, and ions. The amino acid chain was pulled through a nanopore using an external force.

Results & Impact

We showed that a nanopore in an ultrathin MoS_2 membrane can be used to detect and identify all of the 20 standard amino acids in the translocating proteins. The high-precision, single-base resolution and fast biomolecular sequencing using nanopore technology can lead to the fabrication of inexpensive personal healthcare devices that will help provide targeted healthcare. This will enable predictive and personalized medicine and will mark a significant leap forward for clinical genomics and proteomics.

Why Blue Waters

We performed 4,103 molecular dynamics (MD) simulations of systems of up to 50,000 atoms and obtained 66 μs of molecular trajectories. Such expensive computations would not have been possible without a petascale supercomputer such as Blue Waters. LAMMPS, the MD package we use in our simulations, scales almost linearly with the number of cores on Blue Waters.

Allocation: Illinois/125 Knh

PI: Tandy Warnow

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

IMPROVING ACCURACY AND SCALABILITY FOR CORE BIOINFORMATICS ANALYSES

Comparison of multiple sequence alignment methods on simulated (left) and biological (right) datasets. BAli-Phy (red dot) has the best accuracy of all methods on the simulated data, but is only good for Modeler score (precision) and not particularly good for SP score (recall) on biological datasets. Figure from https://doi.org/10.1093/sysbio/syy068.

Research Challenge

This project focused on developing new computational methods with high accuracy to address three fundamental problems in biomolecular sequence analysis (metagenomics, phylogenomics, and proteomics) where prior methods failed to provide adequate accuracy.

Methods & Codes

Performed two major studies regarding protein sequence alignment:

- Compared BAli-Phy (Redelings and Suchard 2006, 2007), the leading statistical method for co-estimation of alignments and trees, to leading protein sequence alignment methods on biological and simulated data.
- Evaluated the impact of integrating the best-performing methods from this first study into PASTA and UPP (divide-and-conquer frameworks).

Why Blue Waters

This study used 230 CPU-years for the BAli-Phy analyses alone and would not have been feasible on other computational systems available to the project team.

Results & Impact

BAli-Phy has outstanding accuracy on simulated data sets, see Figure 1, Left, and incorporating BAli-Phy into PASTA and UPP enables it to scale to 10,000 sequences (Nute and Warnow, BMC Genomics 2016). However, when evaluated on biological protein data sets, BAli-Phy has much poorer precision and recall than the top alignment methods; see Figure 1, Right. Possible explanations for this discrepancy in performance include: model misspecification, errors in the biological benchmarks, and differences between structural and evolutionary alignments.

Allocation: Illinois/169 Knh

PI: Zaida Luthey-Schulten

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

AN EFFICIENT HYBRID STOCHASTIC-DETERMINISTIC SIMULATION TECHNIQUE FOR LIVING CELLS
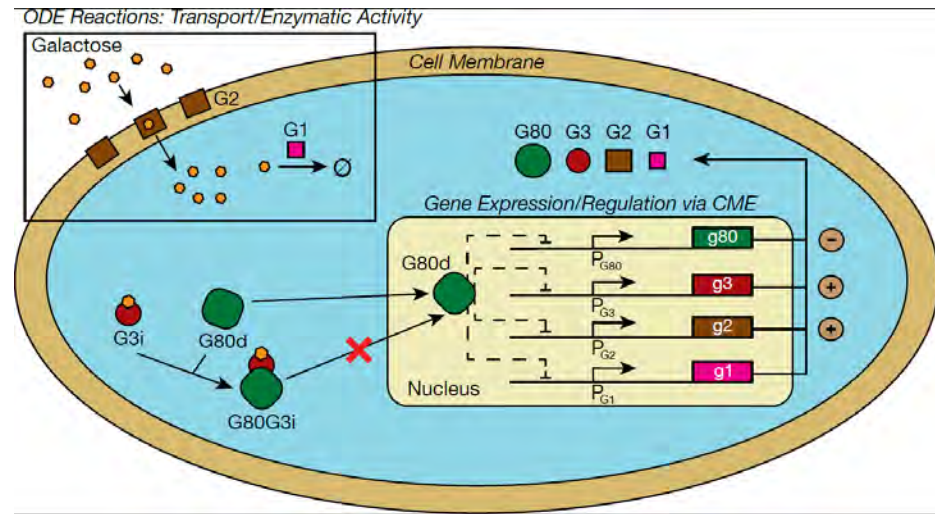
Research Challenge

The Chemical Master Equation and the Reaction–Diffusion Master Equation are descriptions of cellular processes where the system is considered to follow a Markov jump process on the state space of particle numbers in time. Gillespie's Stochastic Simulation Algorithm provides a method for obtaining unbiased realizations of these Markov processes. This algorithm is limited by the fact that reaction events are accounted for explicitly, making simulations of highly reactive systems computationally expensive.

Schematic model of the galactose switch in yeast—The reactions depicted in the boxed area are simulated deterministically via an ODE solver, while those outside this region are simulated stochastically using the SSA.

Methods & Codes

At each time step the LSODA solver is updated with the species counts from the stochastic regime simulated with SSA, and then takes adaptive time steps to evolve the high particle number species through time in the deterministic regime. Then the stochastic rates involving low particle number species interacting with high particle number species are updated with the counts found by the ODE solver. The hybrid algorithm also updates species counts generated from reactions in the CME regime to the ODE regime at this time.

Results & Impact

Simulations enabled by this type of hybrid algorithm will allow researchers to study larger and more detailed systems, capturing the effects of reactions involving high particle count species such as metabolites, which have a crucial role in systems such as the genetic switch studied in this work. This hybrid approach was used to perform a spatially resolved RDME–ODE study of the galactose switch system, experiencing similar speedup to that seen in the CME implementation.

Why Blue Waters

Blue Waters was essential to generate thousands of replicate hybrid simulations over the simulation time of 750 minutes and a range of concentrations. Only then did the study have sufficient data to make the results statistically reliable and to determine the optimal communication time.

IMPROVING NWCHEM SCALABILITY USING THE DATASPACES FRAMEWORK

Research Challenge

The present computational challenge is handling a combinatorial explosion in problem size that accompanies the task of identifying the global minimum in parameter space. The solution is to employ parallelization; however the frequent read / write operations of the large number of parallel processes can overload the filesystem. The aggregate distributed array holds billions of records for an average parameter optimization problem. Sorting an array of such size to determine the promising parameter sets for further analysis represents a practically unsolvable problem that requires a creative solution. Another traditional problem encountered in parameter optimization efforts is the limited availability of training data. Extracting unique molecular clusters from high-resolution X-ray crystals offers a powerful solution but requires the use of computationally very expensive electronic structure computations to generate the corresponding data set.

Methods & Codes

* Target data for parameter optimization is generated by a set of in-house scripts.

* Electronic structure computations employ previously optimized CCSD(T) method in the NWChem package.

* Developed a scalable tool to optimize Lennard-Jones parameters in the AMBER (Assisted Model Building with Energy Refinement) force field.

Results & Impact

This project introduces a procedure for systematic improvement of classical force field by determining the global minimum in the parameter space for an expandable set of the training data. The beneficiary of the optimized parameter set is the entire molecular dynamics community. The developed fractional parallel sorting procedure drastically reduces time spent in sorting as well as the required RAM per node.

Why Blue Waters

With its fast interconnect and large memory per core, Blue Waters is unique in its ability to conduct CCSD(T) computations of molecular systems encountering a thousand basis functions, which is vital for the success of the developed parameter optimization procedure. Since the parameter optimization procedure is extremely resource demanding, the availability of large numbers of nodes is essential for the exhaustive exploration of parameter space.

THE FREE ENERGY LANDSCAPES GOVERNING THE FUNCTION OF COMPLEX BIOMOLECULAR MACHINES
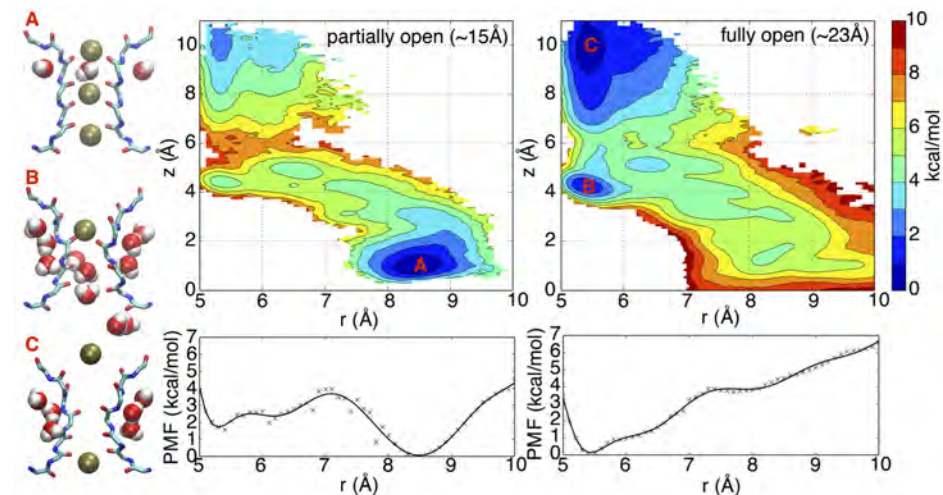
Research Challenge

Proteins such as ion channels, transporters, and pumps play an essential role in controlling the bidirectional flow of material and information across the cell membrane, enabling the cell to accomplish complex tasks. This work investigated the molecular mechanism of the Na+/K+–ATPase ionic pump, as well as the key factors controlling the stability of the C-type inactivated state of the KcsA K+ channel, a nonconducting state of K+ channels with great physiological implications.

Methods & Codes

The timescale limit of molecular dynamics (MD) simulation is overcome by calculating the free energy landscape that governs the key functional motions within a subspace of predetermined order parameters using umbrella sampling (US) Hamiltonian-tempering replica exchange MD (US/H–REMD) simulations. These computations employ the highly parallel molecular dynamics code NAMD extended to treat multiple-copy algorithms.

Why Blue Waters

Potential of mean force to assess the conformational preferences of the selectivity filter with partially and fully open intracellular gate in K+ channel KcsA. The horizontal and vertical reaction coordinates, respectively, describe the width of the selectivity filter (r) and the position of the external K+ ion along Z axis (z).

Li J, Ostmeyer J, Cuello LG, Perozo E, Roux B.

Rapid constriction of the selectivity filter underlies C-type inactivation in the KcsA potassium channel.

J Gen Physiol. 2018 Oct 1;150(10):1408-1420.

Results & Impact

The team found that rapid constriction of the KcsA selectivity filter occurs within 1-2 microseconds when the intracellular activation gate is fully open but not when the gate is closed or partially open. These results imply that the observed kinetics underlying activation/inactivation gating reflect a rapid conductive-to-constricted transition of the selectivity filter that is allosterically controlled by the slow opening of the intracellular gate.

US/H-REMD simulations are scalable to thousands of CPUs; these multidimensional calculations on systems that range in size from 66,104 atoms to 291,148 atoms require thousands of nodes, placing these computations at the forefront of what is possible now, thanks to leadership computers such as Blue Waters.

STRUCTURE AND FUNCTION OF BACTERIAL CHEMOSENSORY ARRAYS

Research Challenge

Individual cells direct their movement in response to chemicals in their environment.

This *chemotaxis* is achieved through an extended, highly-ordered array of chemical receptors.

We do not understand how mechanical elements in the array combine to process chemical signals.

The array structure is the same in all species, making it a potent target for new antibiotics.

All-atom model of a complete bacterial chemosensory array (20 million atoms). Environmental ligands are sensed by binding to chemoreceptors (red), which then transmit signals across the cell's inner membrane to cooperatively regulate the activity of the CheA kinase (blue) with the help of the adaptor protein CheW (green).

Methods & Codes

Cryo-electron microscopy images of the full array were combined with existing X-ray crystallographic structures of individual protein components to create an atomic-resolution model that can be simulated on Blue Waters using the highly scalable molecular dynamics program NAMD.

Results & Impact

Simulations have advanced understanding of the mechanics of chemotactic activation, with specific predictions validated by experiment in live cells.

The development of an all-atom model of a complete bacterial chemosensory array has revealed critical signaling-related interactions among receptors.

Simulation of a complete all-atom model enables tests of existing hypotheses, and drives the generation of wholly new hypotheses.

Why Blue Waters

The required simulations require sustained access to thousands of tightly coupled processors, and benefit greatly from GPU acceleration.

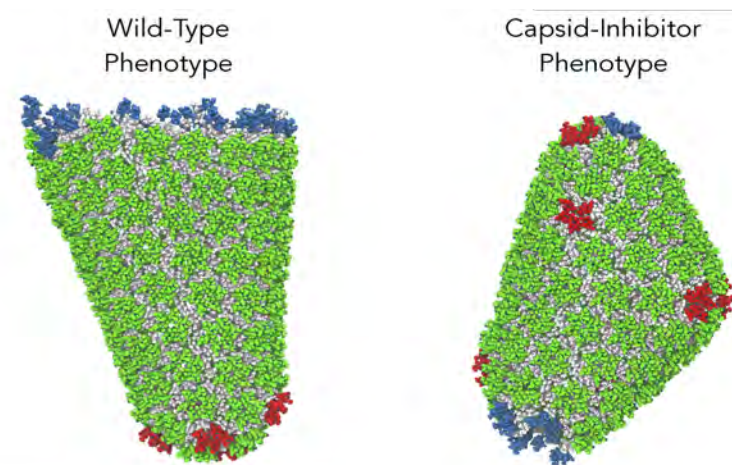Comparison of example end-point structures from wild-type (left) and capsid-inhibited (right) assembly pathways; the N-terminal domains of edge-case dimers, hexamers, and pentamers are shown in blue, green, and red, respectively. In the latter case, accelerated pentameric assembly increases structural polymorphism with notable curvature variability and incomplete enclosures.

LARGE SCALE COURSE-GRAINED SIMULATIONS OF HIV-1: NEW THERAPEUTIC TARGETS

Research Challenge

A key step in the proliferation of HIV-1 is the maturation of viral particles that have been released from infected cells. Proteins within these particles are "activated" through proteolytic cleavage and subsequently assemble into conical cores that house the viral genome. Hence, one viable therapeutic strategy is drugs that prevent proteolytic activity; e.g., protease inhibitors or maturation inhibitors. Recently, a promising class of antiretroviral drugs (GS-CA1 from Gilead Sciences) has been proposed that instead targets the capsid protein directly, but its mechanism of action remains unclear.

Methods & Codes

The team leveraged its previously developed CGMD software and coarse-grained molecular models, which are based on experimental data. These computationally efficient models enable simulations at time- and length-scales that are otherwise inaccessible, allowing simulation of capsid self-assembly.

Results & Impact

Even at small concentrations of drugs, which over-stabilize protein assembly intermediates, a plethora of non-ideal capsid assembly pathways emerge. As a result, the population of canonical and infectious capsid cores is dramatically reduced. The mechanism of action is notably different from classic antiretrovirals; whereas most drugs aim to disrupt or weaken certain interactions, the current class of drugs aims to strengthen interactions.

Why Blue Waters

The Blue Waters combination of leadership-class compute capabilities with cutting-edge network hardware allowed the team to successfully investigate a system of significant biomedical interest. The large-scale simulations were performed using techniques developed for previous work on the Blue Waters platform, while existing relationships with Blue Waters technical project staff greatly assisted in their deployment.

Allocation: NSF PRAC/350 Knh

PI: Ilias Tagkopoulos

University of California, Davis

Biology, Chemistry & Health

Schematic diagram of the knowledge base construction and novel hypothesis generation framework.

AUTOMATIC KNOWLEDGE BASE CONSTRUCTION AND HYPOTHESIS GENERATION: ANTIBIOTIC RESISTANCE MECHANISM FOR *ESCHERICHIA COLI*

Research Challenge

Antibiotic resistance is is one of the leading threats to global health and food security and its direct consequences include longer hospital stays, higher medical costs, and increased mortality rates. Various factors have been preventing the development of machine-learnable knowledge bases that could help better understand and fight antibiotic resistance.

Methods & Codes

The team built a generic framework that constructs an inconsistency-free knowledge graph based on the multilayered machine-learning method and then trains the hypothesis generator using this graph. The team optimized the code to best utilize its principles for the task.

Results & Impact

The team built an extensive knowledge base for *E. coli* antibiotic resistance mechanisms that are specifically structured for efficient machine learning. The knowledge base incorporates a total of one million triples from 10 sources where the five distinctive durations of antibiotic exposure exist, ranging from 30 minutes to 7 days. The proposed framework can accelerate knowledge discovery by automating the process of generation of novel hypotheses from knowledge bases.

Why Blue Waters

The amount of computation and data processing necessary for the project makes Blue Waters critical to the success of this project. Project staff helped the team utilize Blue Waters' computational resources efficiently.

DATA-DRIVEN, BIOLOGICALLY CONSTRAINED COMPUTATIONAL MODEL OF THE HIPPOCAMPAL NETWORK AT SCALE

Research Challenge

Elucidating the mechanisms of sharp-wave ripples (SWRs), which are oscillatory events in the hippocampus required for memory consolidation and subsequent recall. The computational projects aim to construct full-scale, biophysically detailed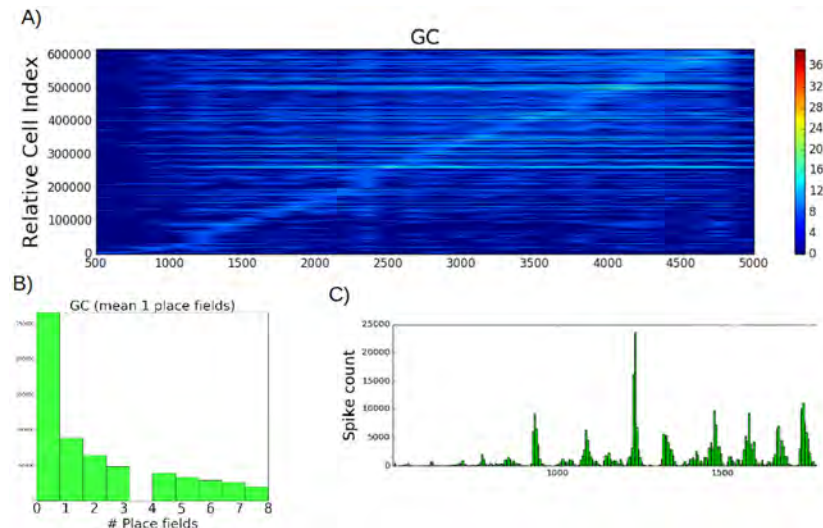 computational models of the three major neuronal circuits within the mammalian hippocampus. These models will help provide insight into the dynamical properties of hippocampal networks that produce the SWR-specific oscillatory patterns.

(A) Firing rate map of granule cells, ordered by peak firing rate. (B) Rhythmic population activity of the major cell types in the network. (C) Consistent with experimental observations, the granule cells in the model have a mean 1 place field and a log-normal distribution of number of place fields.

Methods & Codes

- NEURON as the principal simulation environment.
- Parallel I/O software substrate based on HDF5, creating a parallel computational infrastructure for data management.
- New optimization algorithm, "population annealing," combining aspects of simulated annealing and genetic algorithms to achieve multiobjective optimization in parallel with fewer overall function evaluations.

Results & Impact

- Provides detailed insight on how information propagates in the brain and how the diverse oscillatory dynamics of brain networks emerge. Scientists have not been able to investigate brain function concurrently on multiple organizational levels with reduced-scale computational models.
- The software infrastructure for large-scale neuronal modeling offers a pioneering methodology for rapidly developing, validating, and measuring the information-processing capabilities of realistic, biophysically detailed networks of millions of neurons.

Why Blue Waters

This research requires the simulation of behaviorally relevant brain activity on the scale of tens of seconds. The first year of using Blue Waters, 10-second simulations of models took 14 to 19 hours to run on 1,024 Blue Waters CPU nodes. The PAID program and the technical expertise of Blue Waters' staff helped reduce runtime to three or six hours on 1,024 or 2,048 nodes, respectively. Only Blue Waters permits longer simulations of the eventual combined hippocampal model. The important discoveries revealed by these simulations underscore the essential need for public petascale computing resources such as Blue Waters.

Allocation: NSF PRAC/5750 Knh

PI: Ken Dill

Stony Brook University

Biology, Chemistry & Health

COMBINING PHYSICS AND SUPERCOMPUTERS TO PREDICT PROTEIN STRUCTURES FROM SEQUENCES

CASP competition results produced with MELD on Blue Waters. On top are three protein structures predicted blindly (blue ribbons), along with the experimental structures (red ribbons). Below are the results of all groups in CASP (grey lines) and our prediction (blue line). For T0816, MELD was the only method to provide high-accuracy structures.
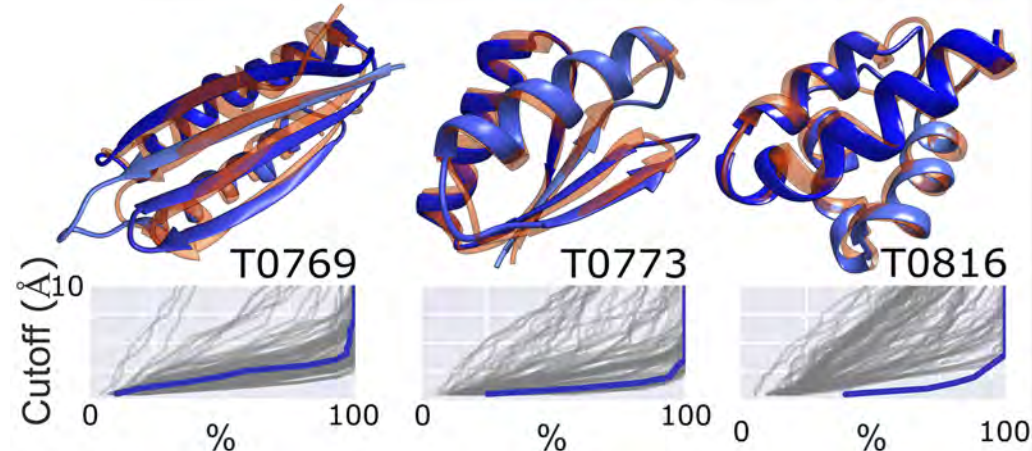
Research Challenge

Proteins consist of a sequence of amino acids that folds into a 3D shape. This shape allows the protein to do its work inside our bodies. Proteins are drug targets, so developing new drugs requires knowing what these proteins look like. However, determining the structures experimentally is time-consuming, expensive, and not always possible. The team uses physics and supercomputers to reduce the time and cost of determining what proteins look like, atom by atom. This has been a grand challenge in computational biology for the last 50 years.

Methods & Codes

The team's MELD (Modeling Employing Limited Data) method is implemented as a plugin to the OpenMM molecular dynamics (MD) package. MELD consists of a replica exchange MD protocol that incorporates biases from experiment, general knowledge, or bioinformatics. What is unique about MELD is that the information is expected to be unreliable. Hence, rather than enforcing all of it, only a fraction is imposed, which changes deterministically at every timestep.

Results & Impact

MELD is the only atomistic physics-based method that is fast enough to compete in the CASP worldwide protein structure prediction competition. MELD can predict "unthreadable" structures that have failed with other state-of-the-art methods because they have no homology to other known proteins. States predicted by MELD are being used to predict the pathways that proteins fold by or use to transition between different states.

Why Blue Waters

Blue Waters is the only system in the United States with enough GPUs for the team to compete in CASP, and the only one allowing many jobs requiring at least 30 GPUs each to run long enough simulations for us. Staff have been instrumental in improving the efficiency of running jobs during the CASP competition.

Allocation: NSF PRAC/12000 Knh
PI: Thomas Cheatham
University of Utah

Biology, Chemistry & Health

USING LARGE ENSEMBLES OF AMBER MOLECULAR DYNAMICS SIMULATIONS TO REPRODUCIBLY AND RELIABLY CONVERGE THE CONFORMATIONAL DISTRIBUTIONS OF NUCLEIC ACIDS

Research Challenge

Molecular dynamics simulations can help us understand biomolecular function, biomolecular assemblies, ligand–receptor interactions, and biomolecular folding.

Key issues include being able to sample conformational space efficiently and effectively and also to accurately model the molecular interactions with appropriate biomolecular force fields.

The image shows an overlay of six average structures of the Drew–Dickerson DNA dodecamer $(CGCGAATTCGCG)_2$ from a 15-microsecond simulation (using the 5 to 10 μs region, all-atom RMS fit) using six different salt combinations: LiCl, LiBr, NaCl, NaBr, KCl, and KBr, each of 200mM concentration.

Methods & Codes

The AMBER molecular dynamics engine PMEMD is well-optimized in CUDA for high performance on GPUs. Multidimensional replica exchange molecular dynamics (M–REMD) reproducibly converges the conformational ensembles of various biomolecular systems. The newly parallelized CPPTRAJ trajectory analysis code enables efficient processing of the output of these simulations.

Results & Impact

M–REMD methods are demonstrated to reliably and reproducibly converge the conformational ensembles of DNA helices and RNA dinucleotides, tetranucleotides, and tetraloops. The AMBER biomolecular force fields perform incredibly well for nucleic acid helices, but anomalous conformations observed in RNA simulations show the potential for further optimization.

Why Blue Waters

The thousands of GPUs and well-performing Lustre parallel file system of Blue Waters facilitated our simulation and analysis workflow.

The Blue Waters team helped us overcome problems and facilitated some CUDA, OpenMP, and MPI optimizations of the CPPTRAJ analysis code.

Allocation: NSF PRAC/6510 Knh

PI: Vincenzo Carnevale

Temple University

Biology, Chemistry & Health

MOLECULAR DYNAMICS SIMULATIONS UNVEIL THE MECHANISM OF THE SENSITIVITY TO HEAT OF AN ION CHANNEL

(A) Coupling between N676 and Y671. Top left panel: Y671 orientation and the π-bulge position. Only two conformations are observed: open state (1) and closed state (2). The bottom left panel shows the distance between N676 carboxamide carbon and I672 carbonyl oxygen and the π-bulge position. (B) A change of Y671 orientation results in a displacement of the pore helix.
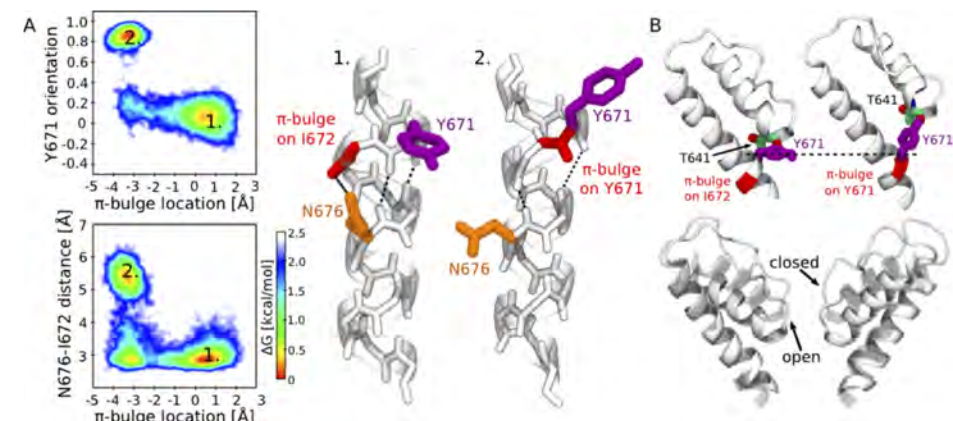
Research Challenge

TRPV1 an extremely promising target for treating chronic pain. Modulating the channel activity for pain control is safer for TRPV1 than for other ion channels involved in propagating painful signals. This is because, despite being relatively widely expressed, TRPV1 channels do not play a crucial role in the heart or central nervous system. However, very little is known about its activation mechanism. This research has discovered a molecular mechanism of activation that involves the rotation of a conserved asparagine in one of the pore lining helices in and out of the pore

Methods & Codes

- Structure of the TRPV1 capsaicin-bound (CAP-bound) state was taken from the Protein Data Bank (PDB). Refined the structure and modeled the missing residues using Rosetta.
- Long time-scale molecular dynamics (MD) simulations using the NAMD package and CHARMM36 force field together with enhanced sampling techniques to estimate free energies

Results & Impact

Discovered that TRPV1 activation involves the rotation of an evolutionarily conserved amino acid located in the middle of the pore lining helix S6 (N676). The conformational switch of N676 is correlated with a wet-to-dry transition in four so far unreported peripheral cavities. These cavities offer a new opportunity to modulate the activation of TRPV1 through small molecule binding.

This study provides a microscopic picture of the activation mechanism, which will allow researchers to expand the chemical space and select optimal compounds.

Why Blue Waters

This project investigated a system of approximately 400,000 atoms using molecular dynamics. The time scales involved in the activation process of TRPV1 dictated trajectory lengths on the microsecond range. The extensive simulation was possible thanks to a massively parallel calculation enabled by the computational capabilities of Blue Waters.

Social Science, Economics, & Humanities

POLICY RESPONSES TO CLIMATE CHANGE

Research Challenge

Leading integrated assessment models assume that climate damages are related to the mean surface temperature of the planet. But climate science shows that when the climate cools or warms, high-latitude regions tend to exaggerate the changes seen at lower latitudes due to spatial heat and moisture transport called polar amplification (PA). The low- (high-) latitude regions would be hotter (colder) if poleward heat transport were absent. PA will accelerate the loss of Arctic sea ice, a potential meltdown of the Greenland and West Antarctica ice sheets, which could cause serious global sea level rise. Furthermore, PA will also affect the likelihood of tipping points, such as the "nearest" three potential tipping points located in the high latitudes of the Northern Hemisphere (Arctic summer sea ice loss, Greenland ice sheet melt, and boreal forest loss).

Methods & Codes

The DIRESCU model was developed to include spatial heat and moisture transport from low latitudes to high latitudes, sea level rise, permafrost thaw, and tipping points. To model spatial heat and moisture transport, the team disaggregates the globe into two regions. To address the tipping points and solve the dynamic stochastic programming problem, they adapt the computational method in DSICE showing high parallel efficiency, with an almost linear speed-up from 30 nodes to 5,000 nodes.

Why Blue Waters

The parallel computational package requires low-latency communications, and this large problem corresponds to solving a Hamilton–Jacobi–Bellman equation with 10 or 11 state variables. Moreover, the team solved the model with many specification cases for analysis. The largest problem solved for DSICE used 3,459 computer nodes and took 7.5 wall-clock hours on Blue Waters.

Results & Impact

A nonlinear certainty equivalent approximation (NLCEQ) method was published to solve efficiently and high-dimensional dynamic stochastic problems with exogenous trends, and it was applied to analyze the effect of climate and technological uncertainty in crop yields on the optimal path of global land use. The work on the DIRESCU model, which studies optimal climate policies under cooperation and various degrees of competition among regions, finds that excluding some of the elements of climate science leads to significant bias in important policy variables such as the social cost of carbon and adaptation.

Allocation: Illinois/400 Knh
PI: Wendy K. Tam Cho
University of Illinois at Urbana-Champaign
Social Science, Economics, & Humanities

Performance comparison: Spatial Path Relinking crossover operator and the basic overlay + expansion crossover. Both take the same input. The difference in fitness between the input and output solution is plotted. A negative difference value indicates fitness improvement.

ENABLING REDISTRICTING REFORM: A COMPUTATIONAL STUDY OF ZONING OPTIMIZATION
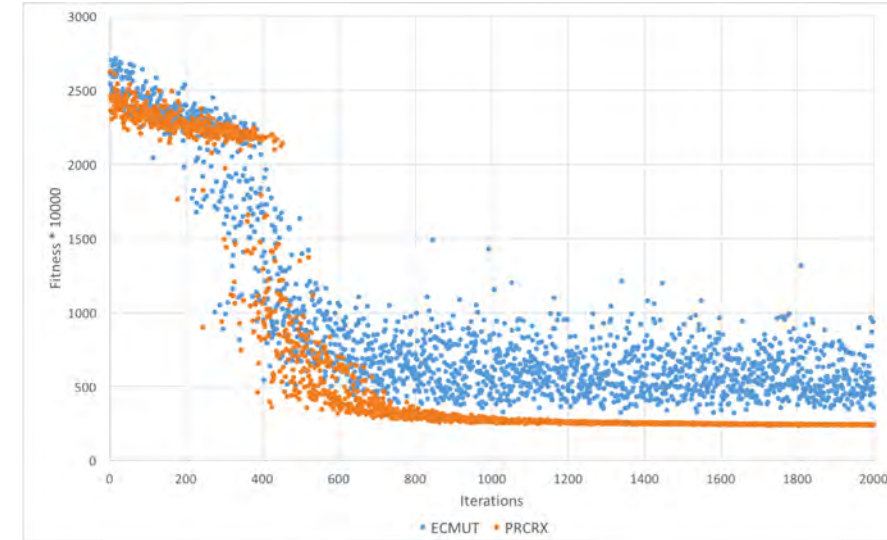
Research Challenge

Political redistricting is intended to provide fair representation in Congress to all communities and interest groups. Gerrymandering occurs when districts are drawn in a manner that discriminates against a partisan or racial group. Due to a lack of sufficient tools to analyze and synthesize redistricting data, the Supreme Court has found it difficult to identify a workable standard by which we might regulate gerrymandering. This team has developed a computational redistricting tool utilizing massively parallel high-performance computing for redistricting optimization and analysis.

Methods & Codes

The algorithm, PEAR, or Parallel Evolutionary Algorithm for Redistricting, uses MPI nonblocking functions for asynchronous migration, and the C SPRNG 2.0 library to provide a unique random number sequence for each MPI process. The design of the evolutionary operators incorporates spatial characteristics to effectively search the solution space. The parallelization of the algorithm maximizes the overlapping of computing and communication at runtime.

Results & Impact

The approach is designed to identify redistricting maps that satisfy a set of user-defined criteria by leveraging and enhancing a scalable Parallel Genetic Algorithm (PGA) library to develop PEAR for the computationally intensive redistricting problem. PEAR provides a novel and powerful tool that harnesses massive computing power to handle spatial characteristics and the associated computational challenges. The project has been the subject of numerous amicus briefs and has been discussed in oral arguments before the Supreme Court.

Why Blue Waters

The PEAR library is designed for extreme-scale redistricting applications. The computational approach implemented in the solution generates a very large number of electoral maps for the quantitative study of redistricting. Identifying quality electoral maps is a very large combinatorial optimization problem that is computationally complex. Generating a large number of feasible and statistically independent maps is possible only on a supercomputer at the scale of Blue Waters.

Graduate Fellows

2014-2015 BW Grad Fellow

Alexandra Jones

University of Illinois at Urbana-Champaign

Geoscience

Vertical cross section of the radiative heating rate through a field of shallow cumulus clouds as calculated by MCBRaT-3D. The top image is the heating rate due to solar radiation; the bottom image is the heating rate due to radiation emitted by the clouds, atmosphere, and surface. The middle image is the net heating rate from the two sources.

HIGH ACCURACY RADIATIVE TRANSFER IN CLOUDY ATMOSPHERES
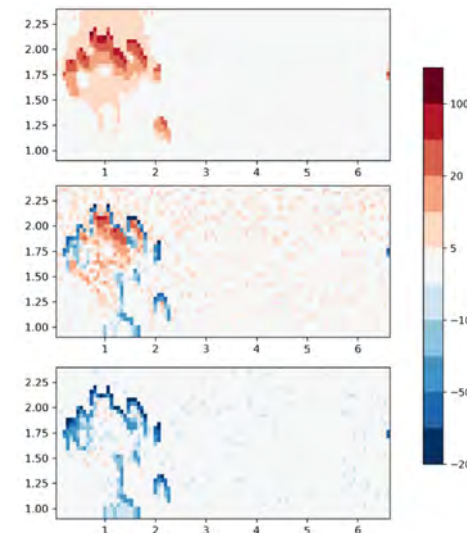
Research Challenge

Because of the important role that clouds play in the atmosphere in redistributing the radiative energy from the sun, Earth, and atmosphere as well as the ubiquity of cloud coverage, it is imperative that we correctly model the interactions between clouds and radiation in order to accurately predict and observe weather and climate. However, modeling of radiative transfer tends to be crude because of the perceived computational expense. Evidence of a bias due to these crude assumptions has been seen in satellite-observed properties as well as modeled cloud properties.

Methods & Codes

A model that treats broadband integration and 3D radiative transfer in a highly accurate and unbiased way is needed to serve as a standard of comparison for similar models and provide accuracy bounds for simpler models and parameterizations attempting to capture 3D effects at lower computational cost. This new approach called MCBRaT-3D uses Monte Carlo methods to capture the 3D transfer of radiation and sample at high resolution the broad range of the electromagnetic spectrum.

Results & Impact

The models, tools, data, and products were made publicly available to the radiative transfer community to aid in faster and more robust progress in addressing scientific questions about the interactions of clouds and realistic radiative transfer. Each of these products has been thoroughly vetted for accuracy and the results of these tests are available for reproduction by other scientists to test these models or their own.

Why Blue Waters

Access to debugging and profiling tools such as DDT and CrayPat allowed the researcher to streamline the development process. Having access to a point of contact on the Science and Engineering Application Support staff helped process through issues and find tailored solutions for development problems resulting in decreased time to discovery. The quick responsiveness of the Blue Waters' staff allowed for limited interruption in progress.

2016-2017 BW Grad Fellow

Andrew Kirby

University of Wyoming

Physics & Engineering

HIGH-FIDELITY BLADE-RESOLVED WIND FARM SIMULATIONS

Instantaneous axial momentum at multiple downstream positions of the National Renewable Energy Laboratory WindPACT–1.5MW wind turbine.

Research Challenge

Predicting wind farm performance represents a complex problem that spans spatial and temporal scales over 10 orders of magnitude from the continental scales that govern wind patterns to the thin boundary layers over the wind turbine blades. The researchers aimed to developed a high-fidelity multiscale modeling methodology to accurately predict the performance of wind turbines by modeling the entire range of these spatial and temporal scales.

Methods & Codes

The software framework uses the Large-Eddy Simulation (LES) approach for prediction of the turbulent flow fields in the off-body region. The team used a multi-mesh framework to provide accurate and efficient prediction capabilities for vortex-dominated wind turbine flow fields. They realized the multi-mesh paradigm by using multiple flow solvers, with each code optimized for the corresponding mesh type.

Results & Impact

These complex aerodynamics, such as flow separation, cannot be captured accurately using lower-fidelity methods. Henceforth, state-of-the-art simulation analysis is now feasible through this computational framework. New pioneering analysis is now accessible, enabling the fundamental understanding of these complex wind turbine wake physics and their interactions. With wake simulated data, improving full wind farm simulations at low cost thereby improving wind farm layout and wind farm efficiency can be derived.

Why Blue Waters

The design of Blue Waters makes it an excellent machine geared toward scientific output rather than just its flop rate. Blue Waters allowed us to perform large-scale wind farm simulations using tens of thousands of compute cores. In addition, the project staff provided excellent insight for optimizing throughput.

Elizabeth Agee

University of Michigan

Geoscience

QUANTIFYING FOREST DROUGHT RESISTANCE
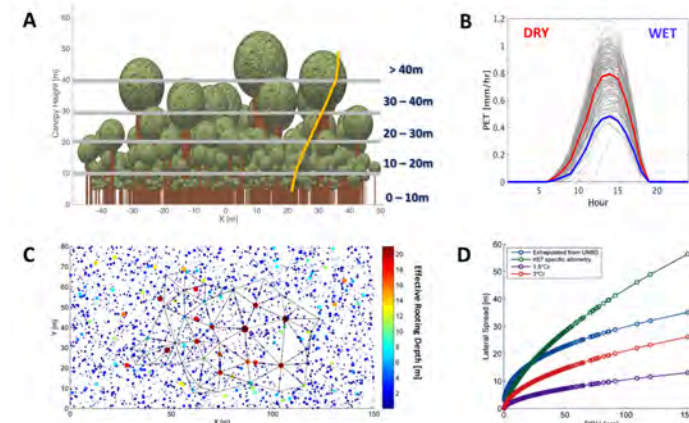
Research Challenge

Over the past two decades, the Amazon Basin region has been hit with multiple drought events that were triggered by strong shifts in sea surface temperature caused by the El Niño–Southern Oscillation. The increased frequency and severity of droughts and their regional consequences have highlighted the potential vulnerability of the Amazon to heat- and drought-induced stress. To adequately capture the response of tropical rainforests to water limitation, mechanistic models that incorporate diverse plant morphology and hydraulic function are needed.

(a) Canopy structure of monitored trees within the Tapajós National Forest. (b) Mean daily potential evapotranspiration for both wet (blue) and dry (red) seasons. (c) Dots represent individual trees in the domain with color indicating rooting depth. (d) Scenarios of lateral spread, which vary with tree diameter at breast height (DBH).

Methods & Codes

Root architectures that represent the structural and spatial distribution of roots have been modeled using the open source RootBox model. Each tree system was assigned hydraulic parameterization based on statistically generated water usage strategies. Root water uptake was coupled with the massively parallel flow and transport model PFLOTRAN, using hybridization techniques. This project is exploring how tree roots contribute to forest drought resilience in areas of the Amazon rainforest.

Results & Impact

Simulations encompass more than 3,000 individuals of varying size and water demands. Preliminary results highlight the contribution of root traits to both individual and community integrity. Analysis is ongoing, but this work represents one of the largest modeling studies of three-dimensional root water uptake ever attempted. Results from this work can enhance next-generation earth system models by identifying key traits for water uptake processes.

Why Blue Waters

Blue Waters is critical to the ongoing success of this project. Simulations of this complexity and scale require the computational power of this system to make meaningful analyses. Not only are the simulation domains complex, multiple simulations are needed to account for system uncertainty. The enormous biodiversity and harsh environmental conditions of tropical forests hinder data collection needed for model parameterization. Scalable, physically based models provide a necessary tool with which to explore modes of uncertainty and help target data collection efforts.

2017-2018 BW Grad Fellow

Erin Teich

University of Michigan

Physics & Engineering

DYNAMICS OF COLLOIDAL GLASS-FORMERS

Relaxation time as a function of system density for two model glass-formers comprised of particles of slightly different polyhedral shapes. Note that the fragilities of these systems, or the slopes of these curves, are dependent on particle shape.
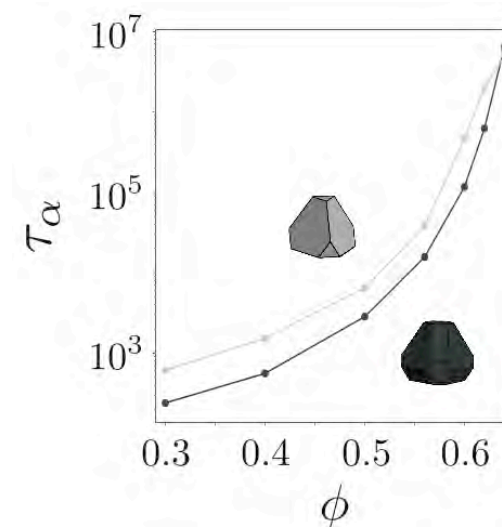
Research Challenge

Glass formation is a well-known outstanding mystery in the physical sciences. Although it has been an aspect of daily life for millennia, it still lacks a canonical thermodynamic explanation. This is due in large part to the significant slowing down of any system as it approaches the glass transition, requiring investigations of glass formation to resolve system dynamics on time scales that vary by orders of magnitude.

Methods & Codes

- Performed hard particle Monte Carlo (HPMC) simulations to model glass and crystal-formers comprised of hard polyhedra contained in the spheric triangle invariant 323 family.
- Used a set of convex polyhedra which is formed by truncating the vertices and edges of a tetrahedron by sets of planes at varying radial distances from the polyhedron center.
- Used the highly parallel simulation package, HOOMD-blue.

Results & Impact

- Simulated systems of 4,096 particles on a single GPU at a variety of densities, and subsequently measured structural and dynamical information. Produced trajectories as large as 3.25 TB per simulation, each containing about 10 million simulation frames.
- Found that colloidal glass-formers exhibited canonical signatures of glassy dynamics, including plateaus indicative of caging in the mean-squared displacement and the real part of the self-intermediate scattering function, and peaks in the non-Gaussian parameter and the self-part of the four-point susceptibility. These peaks indicate dynamical heterogeneity associated with relaxation events.

Why Blue Waters

HOOMD-blue's parallel nature on both CPUs and GPUs has made taking advantage of the computational resources on Blue Waters (and in particular its GPU resources) possible . Blue Waters storage capability also has allowed the generation and storage of tens of TB of raw data on the system for further analysis.

2017-2018 BW Grad Fellow

Ethan Kruse

University of Washington

Space Science

DISCOVERING HUNDREDS OF NEW EXOPLANETS WITH K2
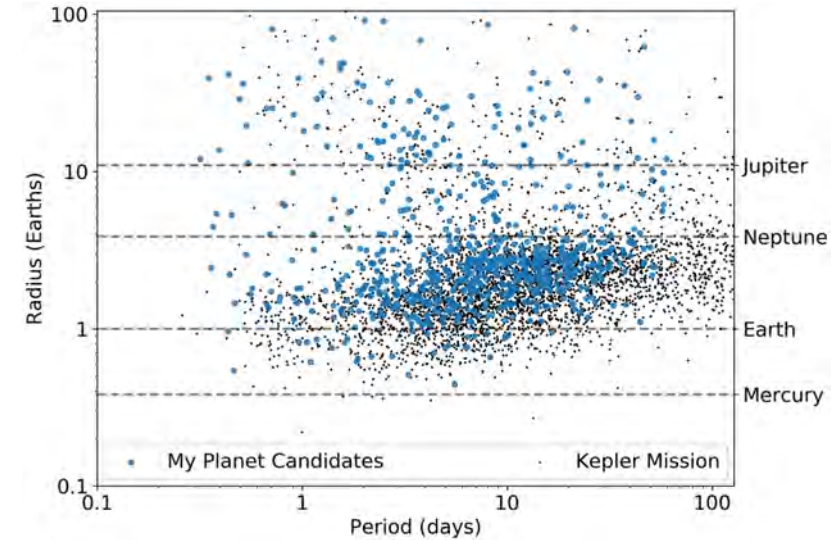
Research Challenge

K2 has succeeded the Kepler mission, and studies more diverse stars. Yet because of the telescope's reduced data quality in its extended mission, searching for planets is not as easy, and there is no official effort to do so. The goal of this project is to develop a pipeline that accounts for the increased noise of K2 data and to search for planets. By finding planets around smaller stars, one canl learn if the radius gap observed around larger stars holds true and get a better handle on what causes it in the first place.

Our K2 exoplanet candidates with those from the original Kepler mission, showing we are finding a similar population of planets. A K2 Campaign is limited to 80 days of observation compared to Kepler's four years, preventing us from finding the longer periods and smaller planets.

Methods & Codes

- Developed a technique that separates the instrumental noise from the astrophysical noise, in a process called pixel-level decorrelation.
- Run a processing pipeline called EVEREST on every star to create light curves with noise at about a factor of four better than the raw light curves: a precision that allows for planets to be found again. These light curves are used on Blue Waters to search for planets.
- Developed the most sensitive and comprehensive planet search pipeline for K2 data using a general-purpose transit search pipeline based on an algorithm called QATS

Results & Impact

- Searched the first two years of K2 data for new exoplanet candidates and discovered over 700.
- Identifying these candidates is just the first step. Adding a large, diverse pool of planet candidates to the sample will enable a suite of follow-up studies.

Why Blue Waters

With over 200,000 stars to search in the first two years of K2 data alone, large computing power is necessary. Access to Blue Waters speeds up development and processing, ensuring that our planet candidates get out in a timely manner for quick follow-up by the community.

2017-2018 BW Grad Fellow

Evan Feinberg

Stanford University

Biology, Chemistry & Health

MACHINE LEARNING HARNESSES MOLECULAR DYNAMICS TO DEVELOP THERAPEUTIC STRATEGIES FOR ALZHEIMER'S DISEASE AND CHRONIC PAIN
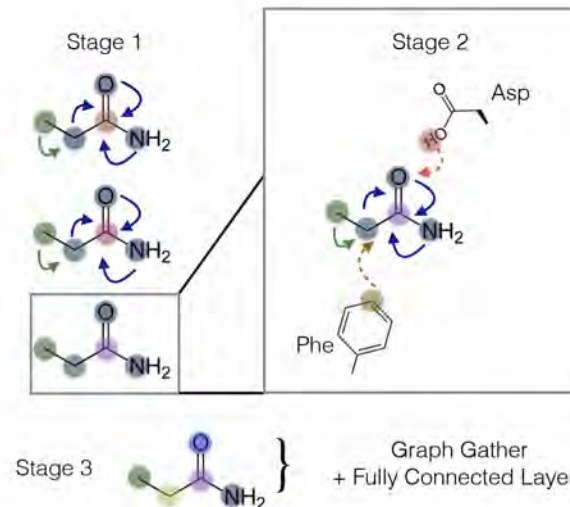
Research Challenge

The arc of drug discovery entails a multiparameter optimization problem spanning vast length scales. The key parameters range from solubility (angstroms) to protein–ligand binding (nanometers) to in vivotoxicity (meters). Through feature learning—instead of feature engineering—deep neural networks promise to outperform both traditional physics-based and knowledge-based machine learning models for predicting molecular properties pertinent to drug discovery. To this end, we developed the PotentialNet family of graph convolutions. These models are designed for and achieve state-of-the-art performance for protein–ligand binding affinity.

Visual depiction of multistaged spatial gated graph neural network. Stage 1 entails graph convolutions over only bonds, which derives new node (atom) feature maps. Stage 2 entails both bond-based and spatial distance-based propagation of information. In the final stage, a graph gather operation is conducted over the ligand atoms.

Methods & Codes

In this project, we generalized a graph convolution to include both intramolecular interactions and noncovalent interactions between different molecules. In particular, we described a staged gated graph neural network, which distinguishes the derivation of differentiable bonded atom types from the propagation of information between different molecules. We implemented the models in PyTorch, a cutting edge deep-learning framework. We trained and evaluated our models on publicly available datasets, including Tox21 for toxicity, ESOL for solubility, and PDBBind for protein–ligand affinity.

Results & Impact

Higher-level interaction "features" were learned through intermediate graph convolutional neural network layers. In light of the continued importance and success of ligand-based methods in drug discovery, we benchmarked PotentialNet on several ligand-based tasks: electronic property (multitask), solubility (single task), and toxicity prediction (multitask). We observed statistically significant performance increases for all three prediction tasks. A potentially step change improvement was observed for the QM8 challenge, which also reinforced the value of the concept of stages that privilege bonded from nonbonded interaction.

Why Blue Waters

The Blue Waters supercomputer, in particular the many GPU nodes, as well as the outstanding staff, were quite important in enabling us to run massively parallel hyperparameter searches to train the optimal deep-learning models for drug discovery tasks.

2016-2017 BW Grad Fellow

Iryna Butsky

University of Washington

Space Science

The primary differences between the galaxy models stem from the distribution of cosmic ray pressure, which is responsible for driving outflows and for altering the temperature and ionization structure of the CGM.

SIMULATING THE CIRCUMGALACTIC MEDIUM: THE NEED FOR ROBUST COSMIC RAY TRANSPORT
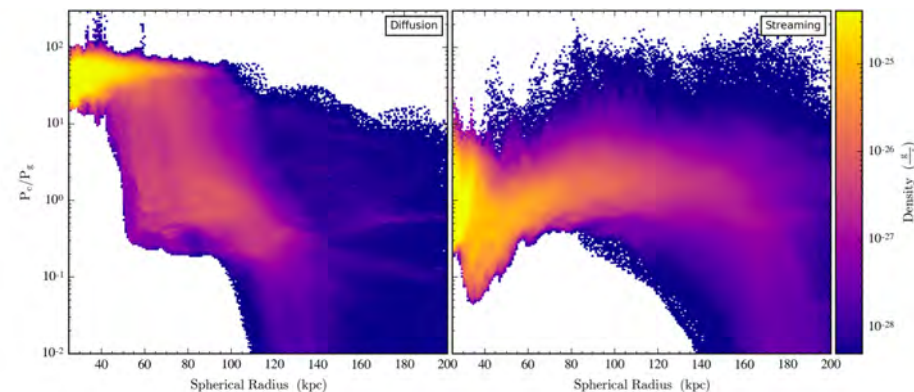
Research Challenge

The majority of galactic baryons reside outside of the galactic disk in the diffuse gas known as the Circumgalactic Medium (CGM). Current simulations excel at reproducing galactic disk properties, but struggle to drive strong galactic winds or match the observed multiphase structure of the CGM with thermal supernova feedback. This project uses a suite of simulated isolated disk galaxies to show that the invoked approximation of nonthermal cosmic ray transport affects the predicted temperature and ionization structure of the CGM and motivates the need for a detailed parameter study.

Methods & Codes

Bulk cosmic ray motion is simulated as a relativistic fluid that is coupled to the thermal gas. Cosmic ray motion relative to the gas is restricted to either diffusion or streaming, depending on the source of turbulence in the magnetic field lines scattering the cosmic rays. The project used cosmic ray diffusion and streaming of Enzo to simulate a suite of isolated Milky Way-type galaxies where supernova feedback injects 10% of its energy as cosmic ray energy.

Results & Impact

Cosmic rays are observed to be a dynamically important component of galaxies and likely play a significant role in shaping the observed ionization structure in the CGM. However, in order for galaxy simulations with cosmic ray physics to hold predictive power, they must first develop robust models of cosmic ray transport.

Why Blue Waters

Realistic simulations of cosmic ray transport must include magnetic fields, which are computationally expensive compared to purely hydrodynamic simulations. These simulations require the use of massively parallel, high-performance supercomputers such as Blue Waters. In addition to Blue Waters' computational resources, the research benefited from the help of the SEAS team, including team member Roland Haas.

2017-2018 BW Grad Fellow

Jennifer Hays

University of Virginia

Biology, Chemistry & Health

$$U_{bias} = \frac{1}{2} k \left(P_{MD}(x) - P_{DEER}(x) \right)^2$$

REFINING THE CONFORMATIONAL ENSEMBLES OF FLEXIBLE PROTEINS USING SIMULATION-GUIDED SPECTROSCOPY
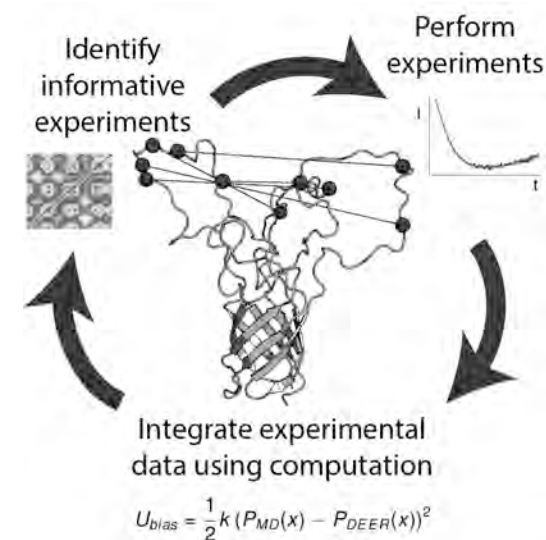
Research Challenge

Many pathogens have proteins that effectively bind to human cells despite being structurally flexible. A computational methodology selects the best experiments to measure these flexible structures and subsequently combine the experimental results in an integrated receptor–ligand model. A systematic approach to refining these flexible receptor–ligand complexes can help elucidate the fundamental physical principles of receptor–ligand binding and promote better drug design for infectious disease.

Iterative refinement of flexible conformational ensemble using simulation-guided spectroscopy. Experimentally-measured distance distributions are integrated into molecular dynamics simulations. Analysis of the resulting hybrid ensemble was analyzed to prospectively identify a set of optimal experiments.

Methods & Codes

Double electron–electron resonance (DEER) spectroscopy is a powerful tool for measuring multiple structures, but DEER experiments are low-throughput, making selection of only the very best, most informative experiments critical. This model-free simulation-based approach for selecting a set of optimal DEER experiments and integrating the resulting data permits estimating the full set of structures at high resolution

Results & Impact

By selecting sets of optimal experiments and incorporating the results into an estimate of the conformational ensemble, experimentalists can now study biological systems that were once prohibitively complex or expensive. This systematic approach to refining flexible receptor–ligand complexes helps elucidate the fundamental physical principles of receptor–ligand binding and promotes better drug design for infectious disease.

Why Blue Waters

Blue Waters has greatly accelerated the time-to-completion of this project. An enormous amount of molecular dynamics sampling is required to capture the full set of structures of a flexible system; that sampling would not have been possible without an exascale system like Blue Waters. Blue Waters provides the computational power to run the required massively parallel MD simulations, scaling to multiple nodes for a single ensemble member, and to many ensemble members.

2015-2016 BW Grad Fellow

Larissa Reames

University of Oklahoma

Geoscience

FACTOR SEPARATION ANALYSIS OF URBAN IMPACTS ON SIMULATED SUPERCELL

Research Challenge

Earth's population is increasingly concentrated in urban areas, with nearly two-thirds of the world's population expected to live in urban areas by 2050. However, interactions between urban areas and synoptically-active convection, such as supercells, remain relatively unexamined. In order to truly understand the nature of these interactions, and thus provide city planning recommendations for the future, it is important to ascertain whether the urban heat island or slower winds induced by increased urban surface roughness result in greater storm modifications.

Methods & Codes

- Weather Research and Forecasting (WRF) model is used for a total of 334 simulations of a supercell thunderstorm to quantify the impacts of a large Great Plains urban area on the evolution and strength of a supercell thunderstorm
- All simulations were run on a 500-m horizontal grid over a 250-km x 250-km grid to properly resolve complex urban structure.
- 120 vertical grid points, with 20 of those points in the lowest 1.5 km above ground were used to resolve the atmospheric boundary layer. More than 29.7 million points over 75,600 timesteps were used for each simulation.

Results & Impact

- Full- and single-physics urban simulations were compared to CTRLE, with the aid of hierarchical clustering analysis (HCA) to form statistically similar groups of simulations; Investigated the effects of the storm having various city-relative path and the storm lifecycle stage during urban interactions;
- Result suggests that urban surface roughness may play a greater role in modifying supercell characteristics than the better-known urban heat island effect.

Why Blue Waters

The large computational and storage requirements of the analysis has made Blue Waters vital to this work. While each simulation was relatively small, the large quantity of of simulations needed to produce significant results required the large computational and data storage capacities of Blue Waters.

2017-2018 BW Grad Fellow

Lauren Foster

Colorado School of Mines

Geoscience

High elevation evergreen trees are sensitive to climate change; these are missed at coarse resolution

Vegetation
- Grasslands
- Evergreen Needleleaf
- Barren
- Deciduous Broadleaf
- Open Shrublands
- Impermeable
- Bare Soil
- Mixed Forest
- Cropland/Natural Vegetation

Reduced streamflow under climate change is due to large evapotranspiration (ET) sensitivity to increased temperature in high elevation, evergreen forests. The effect is only observed in simulations at 100-meter resolution (left plot above).

SCALING RELATIONSHIPS ACROSS MODELING RESOLUTIONS IN MOUNTAIN HEADWATERS: UNDERSTANDING CLIMATE CHANGE IMPACTS ON ROCKY MOUNTAIN HYDROLOGY IN A NUMERICAL MODELING CONTEXT
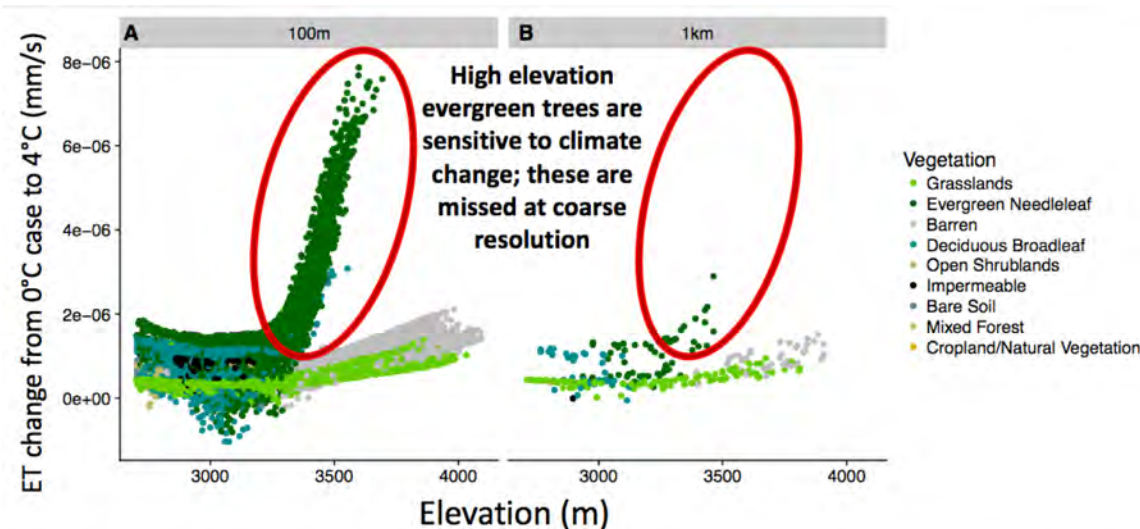
Research Challenge

This project focuses on water stores in the Rocky Mountain headwaters that provide 85% of upper Colorado River streamflow and that have been shown to be very sensitive to a changing climate. Interactions between water and energy fluxes in mountain regions are nonlinear and difficult or even next to impossible to model with simple algorithms. Understanding these interactions is, however, critical to improving predictions of water supplies under climate change.

Methods & Codes

The team modeled a representative headwater catchment in the physically based code Parflow–CLM at 1-km and 100-m resolutions. These models were used to do extensive sensitivity experiments across parameter sets and future climate projections to understand how results and predicated behavior changed across differernt scales.

Results & Impact

The team developed a new method for scaling hydraulic conductivity, a sensitive hydrologic parameter, across modeling resolutions. This method enables calibration of coarse-resolution models that can then be applied at scales where calibration is not possible. The team also demonstrated that existing predictive models may overestimate water supplies in the Colorado River Basin for the coming century by missing non-linear, local-scale processes like increased sensitivity to warming in high-elevation evergreen forests.
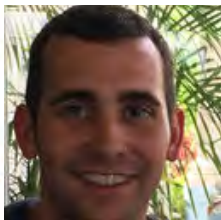
Why Blue Waters

The Blue Waters project provided the graduate support to make this work possible. The consistent interaction with the Blue Waters Point of Contact expanded teams knowledge of HPC systems and applications.

2017-2018 BW Grad Fellow

Matthew Clement

University of Oklahoma

Space Science

HOW THE SOLAR SYSTEM SURVIVED A VIOLENT EPOCH OF DYNAMICAL INSTABILITY
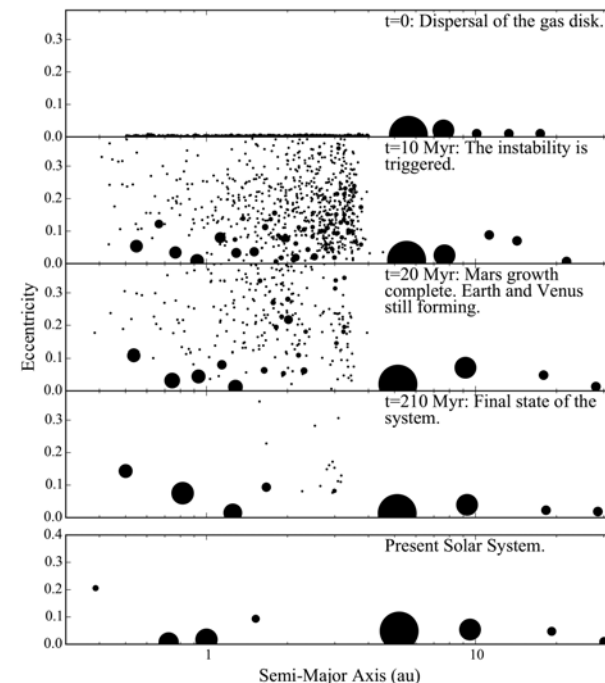
Research Challenge

The research studies how the formation of the solar system's inner planets (Mercury, Venus, Earth and Mars) is influenced by the presence of the outer planets (Jupiter, Saturn, Uranus, and Neptune). The outer planets formed first, while there was still gas present in the primordial solar system. The research challenge is to accurately model the millions of bodies involved the late stages of planet accretion while the giant planets' orbits evolve towards their modern orientation.

Semi-Major Axis vs. orbital eccentricity plot depicting the evolution of a successful system. The size of each point corresponds to the mass of the particle; because Jupiter and Saturn are hundreds of times more massive than the terrestrial planets, separate mass scales are used for the inner and outer planets.

Methods & Codes

- The Mercury6 N-body hybrid integrator (written in Fortran) was used to perform thousands of 200 million year long simulations of the planet formation process
- GENGA, a GPU accelerated code written in CUDA C, was also used to model the complex dynamics in the asteroid belt.

Results & Impact

The team's work offers a simple and elegant explanation for Mars' small size and rapid growth. Isotopic dating indicates Mars reached its present size ten times faster than the Earth and Venus. When the instability occurs, "Mars" is just one of several Mars-sized objects with similar orbits that would have eventually combined into a larger planet. In successful simulations, "Mars" undergoes no further accretion events after the instability. Meanwhile, Earth and Venus continue to grow. Thus, the simulations can explain simultaneously the structure of both the inner and outer solar system.
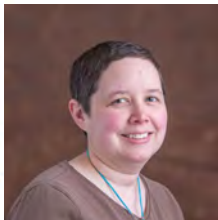
Why Blue Waters

Blue Waters' unique capabilities and highly supportive staff were crucial to the success of this project. The study relied heavily on the use of GPU accelerators on Blue Waters' XK nodes. Having the ability to efficiently run large suites of GPU-accelerated jobs led the team to seek out a Blue Waters allocation. Furthermore, the initial work on Blue Waters has spurred several follow-on projects.

2015-2016 BW Grad Fellow

Maureen Brooks

University of Maryland

Geoscience

Sargassum growth, shown as % maximum growth rate, from model *Sargassum* particles. (a) Control; (b) Inertial experiment. Note patchiness and zonal spreading consistent with entrainment in eddies.

MULTISCALE BIOPHYSICAL INTERACTIONS: INERTIA, GROWTH, AND *SARGASSUM* SEED POPULATIONS

Research Challenge

Sargassum is a macroalgae, or seaweed, that spends its entire life cycle floating on the surface of the Atlantic Ocean. This makes it an ideal model organism for understanding mesoscale biological–physical interactions. It is also of interest because it supports a unique open ocean ecosystem but also has negative consequences for coastal communities when it washes ashore. This study examines how inertial forces influence the pathways *Sargassum* follows across the Atlantic and its access to nutrients vital for growth.
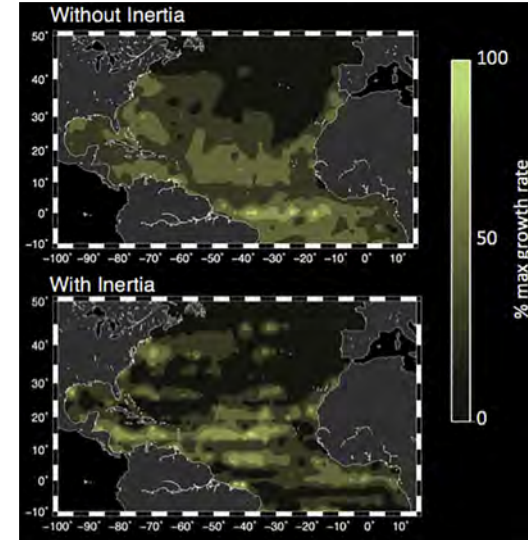
Methods & Codes

A system of four coupled models was developed to simulate *Sargassum* growth and transport.

- A modified version of the Hybrid Coordinate Ocean Model (HYCOM) code
- A biogeochemical model including nutrient chemistry and plankton growth
- A Lagrangian particle model to simulate *Sargassum* raft movement
- A *Sargassum* growth model including reproduction and mortality

Why Blue Waters

The scope of this project is reliant on Blue Waters. This study combines high-resolution ocean circulation with ocean biogeochemistry, Lagrangian particle tracking, and individual-based organism physiology across orders of magnitude of spatial scales. The performance of Blue Waters helped accommodate the high computational cost of running this complex system of models. In addition, the responsiveness and professionalism of the NCSA staff has been of great value in implementing this code on Blue Waters.

Results & Impact

- Results suggest that inertial forces enhance *Sargassum* connectivity between the tropics and the Gulf of Mexico, alter rates of entrainment in eddies, and lead to patchier growth.
- These point to the tropics as the major potential source for inundation events in the Caribbean.
- Understanding this pathway and the oceanographic conditions that drive it is critical to predict and eventually mitigate these costly events.
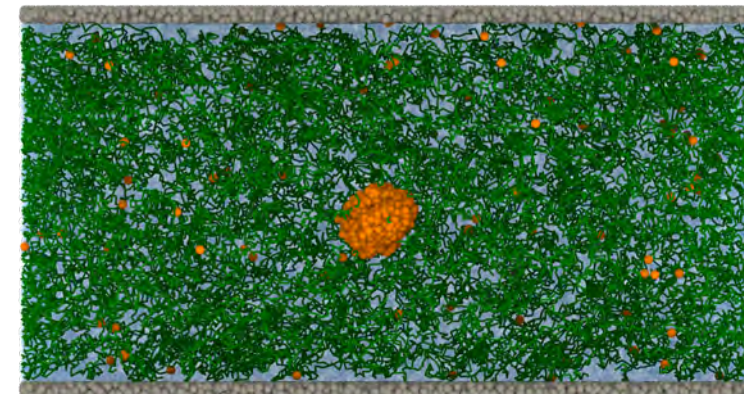
2016-2017 BW Grad Fellow

Michael Howard

University of Texas at Austin

(Fellowship conducted while at Princeton University)

Physics & Engineering

Snapshot of a droplet in a polymer solution under flow simulated using the DPD method. The different DPD particles are: droplet (orange), polymers (green), solvent (blue), and microchannel walls (grey).

DESIGNING MATERIALS IN FLOW: MULTISCALE SIMULATIONS OF COMPLEX FLUID RHEOLOGY

Research Challenge

Complex fluids are multicomponent mixtures that exhibit a rich variety of flow behaviors. For example, corn starch and water ("oobleck") acts like a liquid when pressed slowly but like a solid when struck quickly. Computer simulations allow us to study the microscopic molecular structures and interactions in these mixtures that control such behaviors, but performing the simulations at physically relevant scales presents a considerable challenge. Many complex fluids consist of large solutes suspended in a much smaller solvent such as water, requiring a multiscale approach that simplifies the solvent model while preserving its most important interactions to study complex fluids at relevant length and time scales.

Methods & Codes

Two particle-based mesoscale simulation methods were used to simulate complex fluids in flow: multiparticle collision dynamics (MPCD) and dissipative particle dynamics (DPD). MPCD and DPD significantly accelerate simulations of complex fluids compared to explicit-solvent molecular dynamics models and employ coarse-grained representations of the solvent to reproduce relevant physics. MPCD and DPD models often still require large numbers of particles with simple interactions, naturally inviting a parallel computational approach. The massive parallelism of graphics processing units (GPUs) was leveraged within the HOOMD-blue simulation package to model complex fluids under flow using MPCD and DPD.

Why Blue Waters

Blue Waters is the only system available to us that delivers both the CPU and GPU resources necessary to develop and optimize our MPCD software at scale.

Results & Impact

- The first massively parallel implementation of MPCD for GPUs was released as open-source in HOOMD-blue. Scales efficiently up to **1,024 nodes** on Blue Waters. MPCD performance on XK nodes roughly **3x** faster than XE nodes.

- Research communities can now simulate processes with MPCD at physically relevant length and time scales that would be otherwise inaccessible.

- The migration of droplets in microchannels was simulated using DPD, finding that the addition of polymers to the solvent phase induced migration toward the middle of the channel. This effect may be exploited during oil recovery to enhance the mobility of oil droplets or used to manipulate soft biological materials such as cells in lab-on-a-chip devices.

GENOMIC PERSPECTIVES ON THE AMPHIBIAN TREE OF LIFE

Research Challenge

The relationships among extant amphibians (frogs, salamanders, and caecilians) has been a longstanding debate in phylogenetics, and previous studies have supported any of the three possible topologies relating these three extant orders of amphibians. It is becoming increasingly clear that different regions of the genome can support conflicting phylogenetic hypotheses, and reconciling these discordant gene genealogies is a key problem facing evolutionary biologists today. Information-theoretic approaches have great promise for parsing signal from noise in large phylogenomic data sets. This work seeks to resolve evolutionary relationships among major groups of vertebrates by summarizing phylogenetic information from across the nuclear genome, a computationally intensive endeavor for which access to high-performance computing resources such as Blue Waters is essential.

Methods & Codes

Gene-by-gene tests of constrained topology, comparing maximum likelihood estimates of gene trees for 194 nuclear genes between constraints for the possible interordinal topologies. Akaike information criterion was applied for quantifying the direction and magnitude of support across genes. Used Astral to reconcile gene trees into an estimate of the species tree. This species tree topology was then used in concert with a set of 25 fossil calibrations to estimate divergence times across Amphibia in MCMCTree.

Why Blue Waters

Access to the Blue Waters system provided opportunities to leverage large-scale GPU processors to accelerate the tens of thousands of gene tree estimation analyses needed to implement information-theoretic topology testing.

Results & Impact

- This project demonstrates there is substantial variation in the amphibian genome for which of the three possible topologies relating the three amphibian orders is supported across genes.

- Analyzing hundreds of genes for hundreds of species has traditionally been computationally intractable for empirical data sets, and this study is one of the first to use an information-theoretic framework to address not only the direction of support for phylogenetic hypotheses across the genome but also the magnitude of that support.

2017-2018 BW Grad Fellow

Rachel Kurchin

Massachusetts Institute of Technology

Physics & Engineering

Molecular-orbital-type band structures featuring deep (a) and shallow (b) cation vacancies. Aligning of the orbital energies enables larger band dispersion and hence shallow cation vacancy states.

TOWARD QUANTITATIVE PREDICTION OF DEFECT TOLERANCE IN SEMICONDUCTORS: LEVERAGING HPC AND PHYSICAL INSIGHTS TO DESIGN RADICALLY CHEAPER ENERGY MATERIALS

Research Challenge

The project aims to understand the physical mechanisms underlying defect-tolerant behavior in semiconductors. Silicon, the dominant photovoltaic material, requires an extremely high degree of crystalline perfection in order to perform well, while newly discovered materials can achieve high efficiencies despite the presence of many defects. However, all such new materials prompt concerns surrounding elemental scarcity, toxicity, and stability. Understanding the physics behind the defect-tolerant behavior would enable us to engineer materials that do not suffer from these drawbacks.
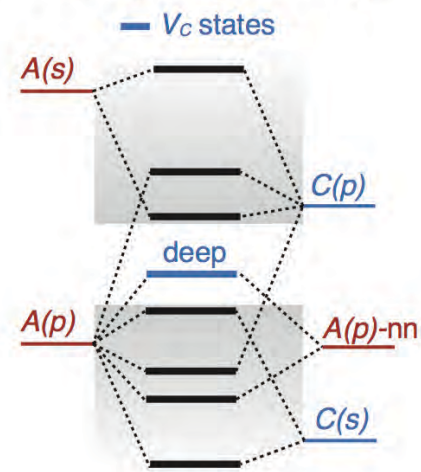
Methods & Codes

The project involves conducting quantum mechanical calculations of crystalline materials using the density functional theory formalism as implemented in the VASP code. The computation delivers formation energies and charge transition levels of various defects of interest in order to reveal the trends in behavior across a wide array of compounds.

Results & Impact

The outcome of this work is the use of computer simulations to inform and guide materials design and discovery — the tasks that, at the moment, are mostly done by trial and error or intuition. The conducted work is part of a coming revolution in HPC-aided precision materials design that is capable of targeting specific properties and functionalities of interest to science as well as society.
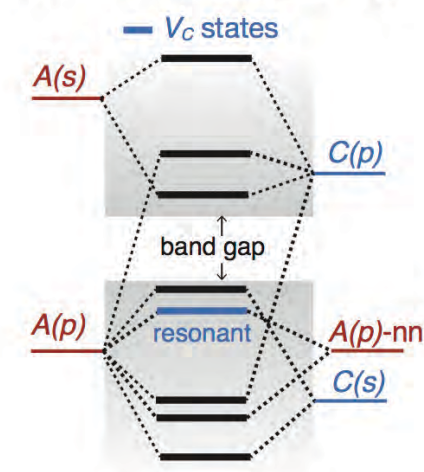
Why Blue Waters

The funding from the Blue Waters graduate fellowship, the community of fellows and of NCSA staff, more broadly, was key to my intellectual independence as well as my becoming a member of the HPC community.

2017-2018 BW Grad Fellow

Rachael Mansbach

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

Snapshots of the assembled morphologies for selected parameter settings (a–h). Side chain beads (SC) are transparent and blue, noncofacial aromatic core beads (BB) are red, and cofacial core beads (A) are green. More ordered morphologies are beneficial.

WIRES WITHIN WIRES: A MULTISCALE MODEL FOR COMPUTATIONAL INVESTIGATION OF BIOELECTRONIC PROTEIN DESIGN
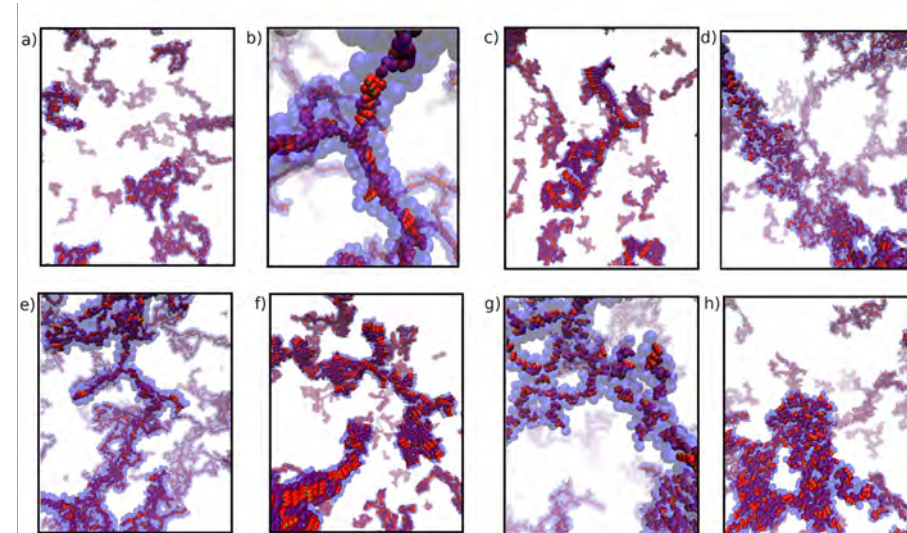
Research Challenge

- Certain proteins containing aromatic groups can assemble under appropriate conditions to form fluorescent mini-"wires" that can be used for electronic applications such as photovoltaic cells, light-emitting diodes, or pH sensors.

- The fluorescent properties of the nanostructures are limited by the extent to which the aromatic cores overlap. To optimize this effect, simulations on long time and length scales must be performed

Methods & Codes

- A simple model of the DXXX series proteins was created in which a single monomer was represented as a set of rigidly constrained beads. HOOMD 2.1.7 was used to conduct Langevin dynamics simulations.

- 10,000 peptide monomers for 660 microseconds with over 60 parameter sets

Results & Impact

- Identified the most salient interaction characteristic controlling the formation of aggregates likely to possess desirable optical properties.

Why Blue Waters

- Simulations would have been prohibitively expensive without access to the computational resources provided by Blue Waters.
- Close support of the project staff was invaluable in enabling us to get up and running quickly.

2016-2017 BW Grad Fellow
Ronald Stenz
University of North Dakota
Geoscience

THE IMPACTS OF HYDROMETEOR CENTRIFUGING ON TORNADO DYNAMICS: IMPROVING THE REALISM OF TORNADO SIMULATIONS

An idealized vortex simulation (left) where an unrealistic buildup of precipitation occurs in the vortex center. The same idealized vortex simulation (right) with centrifuging applied and a physically realistic minimum of precipitation in the vortex center.
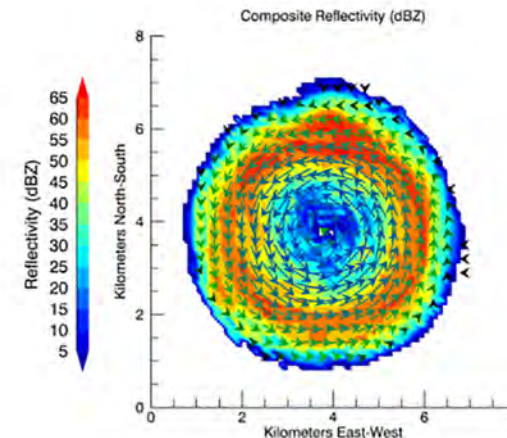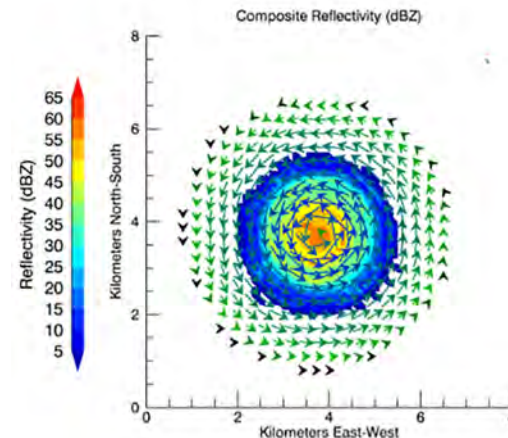
Research Challenge

Current tornado simulations display an unrealistic buildup of precipitation in the vortex center. This project attempts to increase the physical realism of tornado simulations by incorporating precipitation centrifuging, and to quantify its impact.

Methods & Codes

Employed the widely used CM1 (Cloud Model 1) code for simulations.

To quantify the impacts that the inclusion of centrifuging has on tornado dynamics, we first ran simulations without centrifuging and, just prior to the formation of a tornado, saved a checkpoint from which the simulation was continued both with and without centrifuging enabled.

Results & Impact

In simulations without centrifuging, an unrealistic maximum of precipitation develops within the vortex core.

After enabling centrifuging the precipitation in the vortex center is removed and a physically realistic precipitation minimum forms in the vortex center for both the idealized and full-scale tornado simulations

Why Blue Waters

Tornado simulations require thousands of computing cores and produce large amounts of data that must be stored and analyzed.

2017-2018 BW Grad Fellow

Salme Cook

University of New Hampshire

Geoscience

The Piscataqua River-Great Bay estuary at mid-tide—108.8 hours into the model run. Darker blue colors indicate higher water levels than lighter blue colors. The differences in the tidal level from the Atlantic Ocean to the estuary drive large tidal currents that are important for sediment transport.

SEDIMENT TRANSPORT IN ESTUARIES: ESTIMATING BED SHEAR STRESS DISTRIBUTION FROM NUMERICALLY MODELED TIDES IN AN ENERGETIC ESTUARY
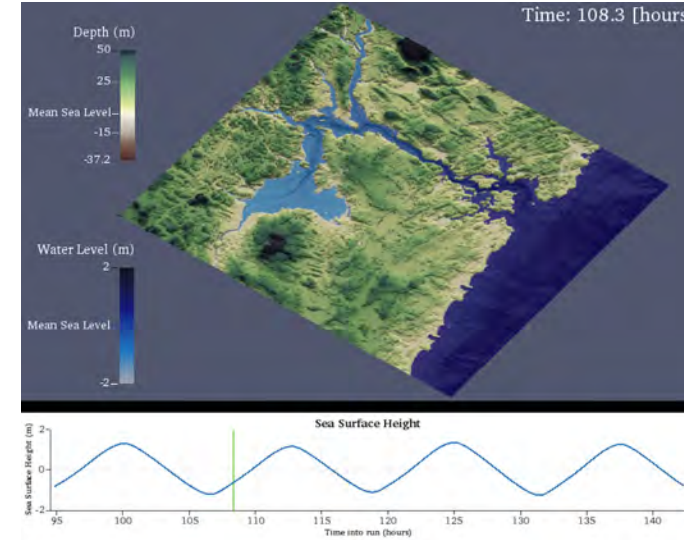
Research Challenge

The rise in land development and associated increases in impervious surface cover have led to a decline in water quality and estuarine health by depositing higher loads of sediment, nutrients, and pollutants. Numerical modeling is a cost-effective way to predict the impact of sediments on estuary-wide nutrient loads, a potentially significant but so far largely neglected source of nutrients.

Methods & Codes

This project used the Regional Ocean Modeling System (ROMS) within the Coupled–Ocean–Atmosphere–Wave–Sediment Transport (COAWST) coupled modeling framework. The model was forced by a combination of tidal and meteorological boundary conditions to reflect real world conditions. Time series analysis and statistical methods were used to determine the best-fit bottom boundary condition for future model runs. The next phase of the project will incorporate waves and sediment transport.

Why Blue Waters

Blue Waters system provided the necessary computational power to test models using a higher-resolution 10-meter grid, which was previously infeasible for this study. Further, the project support staff were an invaluable asset in getting this project up and running on Blue Waters.

Results & Impact

The results suggest that nutrient fluxes derived from sediment resuspension during a typical tidal cycle are potentially significant and should be considered when estimating nutrient loads in estuaries. This is particularly important for estuaries with tidal mudflats, as mud and cohesive sediment tend to trap more nutrients and pollutants. Scientists, land managers, and legislators should incorporate this finding into nutrient load estimates, especially from non-point sources. This is highly relevant for setting project budgets, regulatory limits, and determining best practices for estuarine management.

2015-2016 BW Grad Fellow

Sam Totorica

Stanford University

Space Science

MAGNETIC RECONNECTION IN LASER-DRIVEN PLASMAS: FROM ASTROPHYSICS TO THE LABORATORY IN SILICO
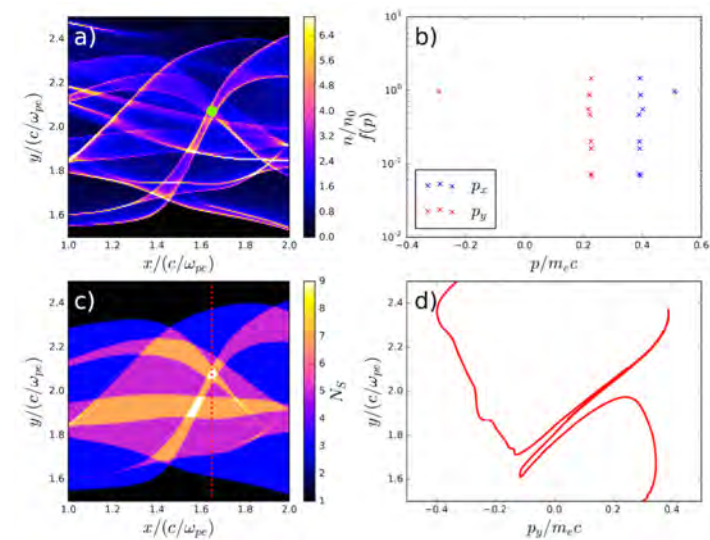
Research Challenge

Magnetic reconnection is a fundamental plasma process that converts magnetic field energy into plasma kinetic energy though the breaking and rearrangement of magnetic field lines. The goal of this project is to use simulations to study particles from reconnection in varied plasma conditions, and in particular to investigate whether laser-driven plasma experiments could be used to study the particle acceleration properties of reconnection in the laboratory.

(a) Charge density for a simulation of the Weibel instability calculated using SIC (simplex-in-cell) and zoomed in on a filament. (b) Momentum distribution at the location indicated by the green circle in (a). (c) Stream number field over the same region and time shown in (a). (d) Transverse momentum along a lineout indicated by the red dotted line in (c).

Methods & Codes

One of the most powerful tools for *ab initio* plasma simulation is the particle-in-cell (PIC) method, which treats the plasma as a collection of discrete simulation particles that interact via electromagnetic forces. The simulations for this project were run using the massively parallel, fully relativistic PIC code OSIRIS, and match the experimental conditions produced by the most energetic laser systems in the world, such as the National Ignition Facility.

Results & Impact

- The team demonstrated that electrons can be accelerated by reconnection with sufficient quantity and energy to be detected in the laboratory.
- Coulomb collisions, included in their simulations, have shown the importance of collisionality on the structure of the reconnection layer
- Simulation results are being used to guide several experimental programs in the United States.
- The team has been developing simplex-in-cell (SIC), a novel method for plasma simulation that, under the right conditions, has been shown to reduce the number of required simulation particles by a factor of 1,000.

Why Blue Waters

This project required the use of large-scale 2D and 3D simulations with sufficient size and resolution to bridge the multiscale physics, from fluid dynamics to the kinetic microscopic processes. These computationally demanding simulations can require billions of simulation particles, and demand the cores, memory, and communication performance available on Blue Waters.

2016-2017 BW Grad Fellow

Sean Seyler

Arizona State University

Biology, Chemistry & Health

Nanojet comparison for liquid argon (1 bar, 80 K): Navier–Stokes (top), linearized 10-moment FHD (middle), full 10-moment FHD (bottom). Blue (red) corresponds to small (large) velocities (0–100 nm/ns); domain measures 500 μm/side (80 cells). The 10-moment models differ only slightly, while Navier–Stokes takes longer to produce jet breakup.

BEYOND NAVIER-STOKES AND MOLECULAR DYNAMICS: UNDERSTANDING BIOMACROMOLECULAR DYNAMICS THROUGH THE DEVELOPMENT OF MULTISCALE HYBRID AND HYDRODYNAMIC SIMULATION
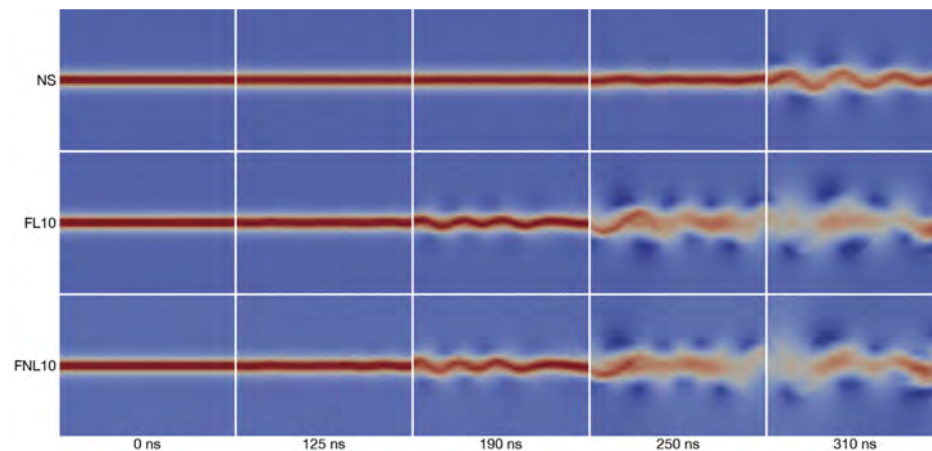
Research Challenge

Proteins are nanomachines that perform mechanical and/or chemical work dynamics spanning femtosecond to millisecond timescales. All-atom molecular dynamics (MD) can fully capture solute–solvent interactions but is currently limited to microsecond timescales—orders of magnitude short of timescales of interest. One means of extending the timescale is the hybrid atomistic–continuum (HAC) method where MD in a subdomain captures atomistic detail while hydrodynamic modelling elsewhere captures solvent dynamics.

Methods & Codes

HERMESHD (Hyperbolic Equations and Relaxation Model for Extended Systems of HydroDynamics) is a numerical fluctuating hydrodynamics (FHD) model to extend the regime of applicability of the Landau–Lifschitz Navier–Stokes equations often used by other HAC methods. HERMESHD can capture nonlinear transport phenomena (e.g., viscoelasticity, thermoacoustic effects, anomalous transport) in liquid water, necessary to explain the emergence of collective phenomena in so-called active matter systems.

Results & Impact

Previously, HERMESHD implemented simplified linearized FHD models with fluctuations. It now implements a nonlinear version of those models, and will soon include the full nonlinear FHD model. Applications of the numerical models to describing molecular motor efficiency as well as collective protein dynamics are currently being explored.

Why Blue Waters

Proper validation of HERMESHD requires extensive testing against both hydrodynamic test problems and computationally-expensive atomistic 3-D MD simulation. The domain-decomposition approach used by HERMESHD depends on the Blue Waters high-speed interconnect to quickly execute such benchmarks. Blue Waters is also ideal for running benchmark MD simulations, which must be sufficiently large to avoid spurious simulation artifacts, to mitigate spurious correlations across periodic boundaries, and to obtain adequate statistical sampling.

2016-2017 BW Grad Fellow

Sherwood Richers

California Institute of Technology

Space Science

The battle of neutrino transport methods. The left side shows the rate at which neutrinos are depositing (red) or removing (blue) energy from a core-collapse supernova using a discrete ordinates method (top) and my Monte Carlo method (bottom). The difference (right) is small, but significant.

GENERAL RELATIVISTIC NEUTRINO RADIATION TRANSPORT: UNDERSTANDING THE DRIVING FORCE BEHIND COSMIC EXPLOSIONS

Research Challenge

Neutrinos ordinarily interact so weakly with matter that experimentalists have great difficulty even detecting them, but they are what drive most stellar explosions. Simulating the behavior and impact of neutrinos remains particularly challenging, since it is a seven-dimensional problem (three spatial dimensions, two independent directions, neutrino energy, and time). The results of this work help us to interpret models and observations and to pave the way for new computational methods that are both accurate and efficient.
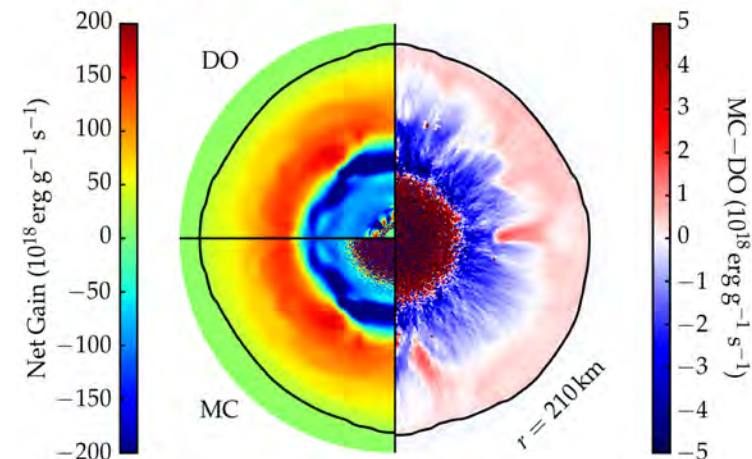
Methods & Codes

- Calculated the steady-state radiation field in core-collapse supernovae and neutron star mergers.
- Used the open-source neutrino interaction library NuLib to calculate the rate that neutrinos interact with the star and other neutrinos.
- Used Sedonu, an open-source, relativistic Monte Carlo neutrino radiation transport code to obtain detailed neutrino distribution information.

Results & Impact

These calculations of Monte Carlo radiation transport are far more accurate than more efficient methods and elucidate where the latter need to be modified to enable high-accuracy simulations without the large expense of Monte Carlo radiation transport calculations. The detailed results pave the way for the next generation of models.

Why Blue Waters

Simulating around 1.3 trillion particles requires: (1) a large amount of memory on each node so each node can fit the entire problem, and (2) many nodes to repeat the problem enough times to get a solution with little noise.

Blue Waters is the prime computer of its time for doing these calculations.

2014-2015 BW Grad Fellow

Varvara Zemskova

University of North Carolina at Chapel Hill

Geoscience

ENERGETIC DYNAMICS OF OCEAN BASIN MODEL WITH ROTATION, SURFACE WIND, AND BUOYANCY FORCING

Instantaneous 3D field of nondimensionalized density for the simulation with realistic buoyancy and wind stress forcing at a time within the statistical steady-state period, showing a range of large- and small-scale eddies.

Research Challenge

The Southern Ocean plays an important role in the global meridional transport of heat and tracers, oxygenation of the bottom waters via the formation of dense Antarctic Bottom Water, and the sequestration of atmospheric carbon.
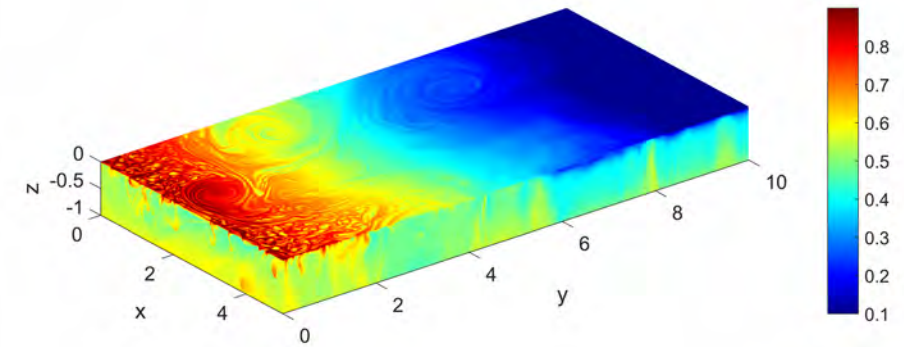
The response of the Southern Ocean circulation to changes in surface wind patterns due to changing climate has been one of the pivotal questions in oceanography. The changes in meridional and vertical transport would affect the rates of carbon sequestration or its release back into the atmosphere as well as the supply of nutrients supporting primary production in surface waters and the ocean food web.

Methods & Codes

Direct numerical simulations (DNS), resolving the smallest dissipative scales, of an idealized rotating rectangular ocean basin, with variable surface density and differing surface wind stress profiles, were performed using the SOMAR code. Each run was analyzed for the circulation stream functions and terms of the energy budget as well as the exchange rates between the kinetic and available potential energy reservoirs.

Why Blue Waters

The direct numerical simulations are highly computationally intensive because of the great spatial resolution required for the small dissipative scales. In addition, these simulations need to be run for 70,000 to 100,000 timesteps, which takes several months, in order to reach a statistical steady state. Both the temporal-average and the deviations from the mean field, meaning that output files (approximately 2.8 GB per file, one per timestep) have to be stored for further analysis.

Results & Impact

This is one of the first DNS studies of an idealized model of the Southern Ocean, resolving all energy scales and dividing the flow into mean and turbulent components. As the wind stress, and subsequently the kinetic energy generation of the system, increases, the increase in the mean flow circulation that is wind-driven is compensated for by dissipation via smaller, transient turbulent eddies, and thus the dissipation of available potential energy is unaffected by the wind stress magnitude. Diapycnal mixing, which supplies oxygen to the bottom waters and is responsible for carbon sequestration, is affected by the surface water density modification due to heating and cooling.

2017-2018 BW Grad Fellow

William Payne

University of Nebraska Medical Center

Physics & Engineering

Simulation of different polymer strands interacting with dye molecules. This process demonstrates drug loading into a polymeric nanoformulation, but the use of dye molecules in place of drugs allows for better experimental evaluation.

COMPUTER-AIDED FORMULATION DESIGN FOR IMPROVED NANOFORMULATION OF ANTICANCER THERAPEUTICS: USING SIMULATION TO IMPROVE THE EFFECTIVENESS OF CANCER DRUGS
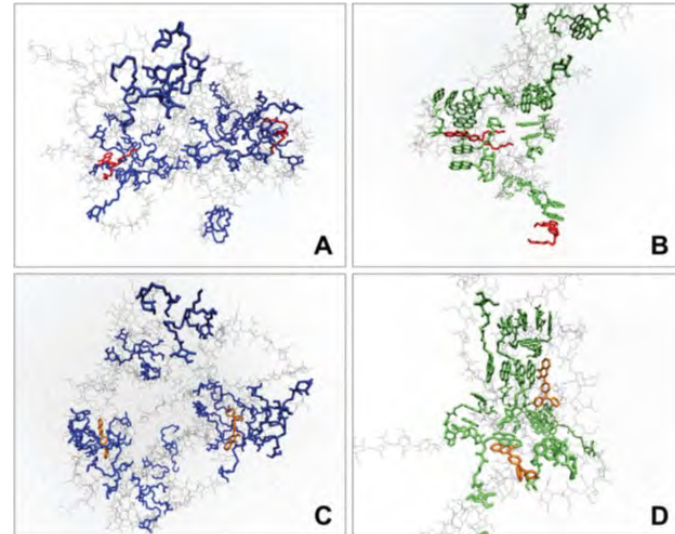
Research Challenge

Typical anticancer drugs produce harmful side effects and can be ineffective. Choosing an ideal drug carrier vehicle is difficult, and there are very few established design processes for the development of new nanoformulations. Development of new nanoformulations usually requires extensive synthesis and experimental evaluation. Ways to guide experiments and to better understand what makes a "good" formulation could dramatically reduce development time and cost.

Methods & Codes

The molecular dynamics programs GROMACS and NAMD were used to model molecular systems across multiple size scales. First, single polymer strands, then multi-strand systems, and, finally, systems including polymer molecules as well as organic dye molecules acting as surrogate drug molecules were modeled. By using dye molecules, simulation results could be compared to experiments, confirming model accuracy and further expanding the information available on important nanoformulation interactions.

Results & Impact

This work was among the first to investigate self-assembled polymeric nanoformulations using simulations and correlating simulations to experimental observations. The simulations have enabled the researchers to focus experimental investigation of new lead formulations as well as to begin working with new polymers in a more effective, methodical design process.
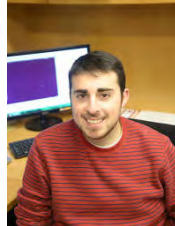
Why Blue Waters

The support, tools, and knowledge provided by the Blue Waters staff dramatically reduced the learning curve to incorporate computational techniques into experimental investigations. The Blue Waters supercomputer permitted more simulations, as job queue times were shorter and Blue Waters performed simulations faster than other available resources.

2017-2018 BW Grad Fellow

Zachary Goldsmith

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

ELECTRONIC STRUCTURE OF NIFE OXYHYDROXIDE ACTIVE SITES: UNDERSTANDING AN EFFECTIVE OXYGEN-EVOLVING CATALYST AT A QUANTUM-MECHANICAL LEVEL

Research Challenge

Facilitating the redox chemistry between H_2O and O_2 is essential to the development of renewable energy technologies. Doing so requires understanding and optimizing efficient and earth-abundant electrocatalysts. NiFe oxyhydroxide is among the most active oxygen-evolving electrocatalysts that has high activity of Fe dopants at the edges of the material. This project aims to determine the nature of these active edges using quantum-mechanical simulation techniques to allow experimentalists and computationalists alike to rationally design even better catalysts for the oxygen evolution reaction.
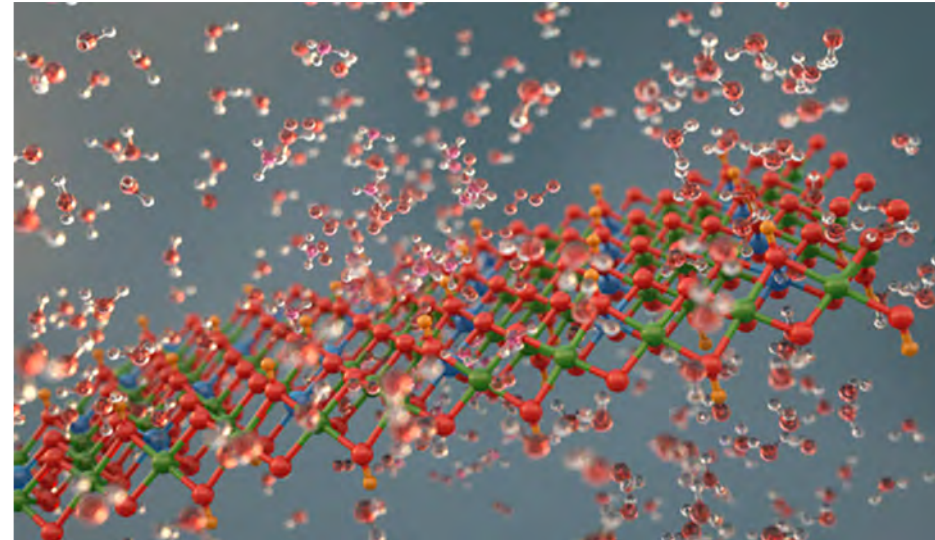

Illustration of a single layer of NiFe oxyhydroxide with 25% Fe doping solvated in water. Ni is shown in green, Fe in blue, O in red, H on the NiFe slab in orange, and H in solution in white.

Methods & Codes

The project involves density–functional theory (DFT) calculations with a hybrid functional for Ni-only and NiFe oxyhydroxide nanowire supercells using Quantum–ESPRESSO code. The model system contains different interior and exterior metal sites. The study probes different degrees of protonation and hydroxylation of Ni and Fe edge sites as a proxy for the different reactive intermediates. Projected density of states around the Fermi energy determines the catalytic activity of studied sites.

Results & Impact

Computation provided improved understanding of the nature of the purportedly active edge sites, particularly in the doped material. For one, there is a ~ 2 eV energetic preference for Fe to dope an exterior site compared to an interior site. The projected density of states shows catalytically active Fe oxide motifs when the Fe dopant resides at the material edge. The chemical nature of the material conduction band minimum determines which sites are most catalytically active.

Why Blue Waters

Blue Waters enabled conducting computationally expensive and highly parallelized hybrid functional DFT calculations to accurately determine electronic structure of Ni and NiFe oxyhydroxide. The modeling of heterogeneous materials interfaces is an important task that can only be done with the frontier scientific computing resources such as Blue Waters.